A NEW THEORY OF COUNTERFACTUAL CONDITIONALS

By

JOHN PEOPLES

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2000

Copyright 2000

by

John Peoples

ACKNOWLEDGMENTS

I owe thanks to a number of people. For their support and encouragement during all my years in graduate school, without which I would never have got through it all, I thank my parents, Richard and Penny Peoples. I would like to thank all the members of my Ph.D. committee: Kirk Ludwig, Greg Ray, John Biro, and Jean Larson. I am particularly indebted to Drs. Ludwig and Ray, who have had more influence on my philosophical development than anyone else. I could not have hoped for two better supervisors. For helpful comments and discussion, I thank Gene Witmer, Robert D'Amico, Kathrin Koslicki, Donald Nute, and the members of the Logic Seminar in the Department of Mathematics at the University of Florida. For help in dealing with all things bureaucratic, I thank the office staff of the University of Florida Philosophy Department: Virginia Dampier and Kitty Powers. Finally, I would like to thank all my graduate student colleagues for their help and support, and particularly Chris Lubbers and Dan Boisvert for being such valuable friends.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ... i

ABSTRACT ... iv

CHAPTERS

1. INTRODUCTION ... 1

2. THE TRADITIONAL DISTINCTION ... 11
   Counterfactuality ... 14
   A Grammatical Distinction? ... 13
   Diagnosis of the Traditional Distinction ... 25
   Summary ... 33

3 LINGUISTIC ANALYSES ... 35
   Early Entailment-Based Theories ... 40
   Later Entailment-Based Theories ... 47
   "No Truth Value" Theories ... 48
   Summary ... 58

4 POSSIBLE WORLDS ANALYSES ... 62
   Material Conditionals and Strict Conditionals ... 66
   Similarity-Based Theories ... 68
   Minimal Change Theories ... 82
   Nute's Analysis ... 89
   Summary ... 93

5 THE SEMANTICS OF TENSE ... 94
   Truth-Theoretic Semantics ... 94
   The Quantificational Treatment of Tense ... 96
   Tense Logic ... 100
   Comparison of the Two Approaches ... 102
   Summary ... 109

6 INDICATIVE CONDITIONALS ... 111
   Jackson, Grice and the Equivalence Thesis ... 111
   Some Objections Answered ... 128
   Summary ... 135

7 A TRUTH-THEORETIC SEMANTICS FOR CONDITIONALS.................. 137
A Semantics for Backshifted Conditionals... 141
Conditionals of the simple past/future-in-the-past combination used to talk about the past..................................... 141
Backshifted conditionals used to talk about the present or future............ 148
Conditionals of the past perfect/future-perfect-in-the-past combination used to talk about the past..................................... 155
Summary.. 165

8 BUSINESS ARISING.. 166

9 CONCLUSION.. 185

REFERENCES... 189

BIOGRAPHICAL SKETCH... 193

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

A NEW THEORY OF COUNTERFACTUAL CONDITIONALS

By

John Peoples

May 2000

Chairman: Gregory Ray
Major Department: Philosophy

In *Fact, Fiction and Forecast*, Nelson Goodman describes the Problem of Counterfactual Conditionals as the problem of defining "the circumstances under which a given counterfactual holds while the opposing [indicative] conditional with the contrary consequent fails to hold." Of course, this definition of the problem assumes the distinction between counterfactual/subjunctive and indicative conditionals that is traditionally made by philosophers and logicians.

In this dissertation, I advance a novel account of the semantics of so-called subjunctive or counterfactual conditionals that denies the traditional distinction between two kinds of conditionals. I note that the conditionals that philosophers and logicians usually refer to as subjunctive or counterfactual are invariably marked by an unusual use of tense. Auxiliary verbs such as 'were', 'had', and 'would' appear in such conditionals where, it seems, they would not normally appear in so-called non-subjunctive cases. I argue that the conditionals that philosophers call subjunctive or counterfactual are a special sort of tensed indicative conditional and require no unusual semantic treatment. I illustrate how these so-called subjunctive conditionals can be treated in a truth-conditional semantics for a natural language. The past-tensed verbs that appear in their protases I take at face value, as about past events and states, and the ‘would’ that almost invariably appears in their apodoses as a relativized future modal (the so-called future-in-the-past). The standard interpretations of such conditionals I treat as the result of the pragmatic implications expected from this interpretation. By denying the traditional distinction, I dissolve the problem of counterfactual conditionals as it is traditionally described, and I provide a framework for more fruitful theorizing about these conditionals.

CHAPTER 1
INTRODUCTION

Traditionally, philosophers have distinguished between two classes of conditionals. First, there are ordinary, so-called indicative ones, and then there is a special class, known as either "subjunctive" or "counterfactual" (these terms often appear to be used interchangeably). The philosophers who make this distinction allege that the semantics (or the truth conditions) of these special conditionals is unlike that of the 'indicative' ones. Most of them say that their truth or falsity depends not only on the actual world, but also on certain other 'possible worlds' or 'hypothetical situations'. Others argue that it depends on special entailment relations not required for ordinary indicatives.

The analysis of so-called counterfactual or subjunctive conditionals has important ramifications for many areas of philosophy. As Nelson Goodman (1955) says, without any means for interpreting these conditionals, we cannot lay claim to any adequate philosophy of science. The problem, as traditionally understood, is to explain the semantics and conversational use of a certain variety of conditional statement. Yet, how do we define exactly which statements we purport to analyze? It is now a commonplace that we would go a long way to solving the problem if we could only get the terminology right, and it is certainly true that confusions and vagaries in terminology plague the literature.

It is curious how so many of those who write about these ‘special’ conditionals never precisely define what they are trying to analyze. Consider, for instance, the following remark of Roderick Chisholm at the beginning of his article, “The Contrary-to-Fact Conditional”:

> [M]any contrary-to-fact conditionals are not expressed in the subjunctive mood and many conditionals which are expressed in this mood are not actually contrary-to-fact, but in the present discussion we may use the labels 'subjunctive conditional' and 'contrary-to-fact' conditional interchangeably. Neither term is adequate, but each has been used in the literature. The essential characteristics of this important type of statement will be more clearly delineated as we proceed. (Chisholm, 1946, p. 9)

Michael Ayers notes that when Chisholm talks about “this important type of statement” he is clearly talking about some third kind of conditional that he believes is neither necessarily counterfactual nor subjunctive. Yet Chisholm fails to fulfill his promise to “more clearly delineate” this class. Ayers contends that this third class “a hazy chimera, an impossible fusion of both [the other two kinds]” (Ayers, 1965, p. 348). I sympathize with Ayers. To use the expressions “subjunctive conditional” and “counterfactual conditional” synonymously is, at the very least, an example of philosophical laziness. At its worst, this looseness of terminology hides the fact that there is no semantically relevant distinction between indicative and subjunctive or counterfactual at all.

Why has this intolerable sloppiness become so widespread? Why do philosophers continue to subscribe to a distinction they cannot define? It is not difficult to see why one might see a need for such a distinction. Let us consider a notorious example, originally due to Ernest Adams (1970):

(1) If Oswald didn't kill Kennedy, somebody else did.
(2) If Oswald hadn't killed Kennedy, somebody else would have.

The second conditional seems to tell us something about a contrary-to-fact situation in which Oswald failed to shoot the President. When confronted with (2), we are prompted to consider various possible conspiracies and plots in order to decide whether it is a conditional we want to accept. An utterance of (1), on the other hand, does not seem to imply anything quite so controversial. An utterance of (1) is plainly acceptable because of the fact that we know that Kennedy was killed, and if Oswald did not do it, there must have been another culprit. Clearly, there is some difference between (1) and (2), and this difference seems to have motivated many philosophers to say that conditionals like (1) and (2) belong in two different semantic categories.

I do not believe that "subjunctive" or "counterfactual" conditionals form a special semantic category distinct from "indicative" conditionals. I argue instead that most of the conditionals that philosophers call "subjunctive" or "counterfactual" are a special sort of tensed indicative[1] conditional and require no unusual semantic treatment. This line of thought is consistent with the view held by many grammarians, for instance, Jespersen (1954-58), that there is very little use of the subjunctive in modern English. I illustrate how these so-called "subjunctive" conditionals can be treated in a truth-conditional semantics for a natural language, along the lines first suggested by Donald Davidson (1984). The past-tensed verbs that appear in their protases I take at face value, as about past events and states, and the 'would' that almost invariably appears in their apodoses as a relativized future modal (the so-called future-in-the-past) (see Chapter 5 for a discussion of relativized past tenses). This treatment results in a number of conditionals that we would not normally assert being found true and some that we would assert being found false. I employ a Gricean argument to explain this phenomenon. I contend that the standard interpretations of so-called "subjunctive" or "counterfactual" conditionals should be treated as the result of the pragmatic implications of their assertions expected from this interpretation. Accordingly, I am able to explain, for instance, why we sometimes use past tense verbs to talk about hypothetical events in the present or future, e.g., 'If I ate that bowl of chili, I'd have terrible indigestion'. On my account, this statement is likely to turn out trivially true in the mouths of most speakers. Therefore, in order to interpret such a statement, we must look at what grounds a speaker has for asserting it under the assumption that she is being conversationally helpful. To oversimplify somewhat, in this case, I claim that we should interpret the speakers use of tense as implicating some causal connection between any event of her eating chili and her having indigestion at some time after.

---

[1] Since I deny the distinction between indicative and subjunctive, my use of 'indicative' here is strictly redundant. However, I continue to use it since I propose to treat so-called 'subjunctive' conditionals in the fashion in which people who buy into the distinction would treat what they call "indicatives."

My proposal is significantly different from the philosophically standard Lewis-Stalnaker possible worlds-based accounts of so-called "subjunctive" conditionals. On my account, such conditionals are not about possible worlds at all. It will also turn out that many (if not most) so-called "subjunctive" conditionals that are found to be false according to possible worlds-based accounts are found to be true on my account. This, of course, will require some explaining. My diagnosis of the thinking behind such accounts is that, although Lewis, Stalnaker *et al* may have unearthed some important intuitions about so-called "subjunctive" conditionals, these intuitions are about pragmatic, not semantic phenomena.

Other philosophers have undoubtedly contemplated the idea of treating so-called "subjunctive" conditionals as indicative conditionals and explaining the differences at the pragmatic level, but they have invariably concluded that it is a dead end. Frank Jackson (1987) is perhaps the most notable example. However, with the notable exception of V. H. Dudman, none of these philosophers takes proper account of the tensed verbs that appear in these conditionals. It will become quickly apparent that tense plays a very significant part in providing the standard import of utterances of these conditionals.

I present a semantics for conditionals in English only. However, if the analysis works for English, we can reasonably expect it to work for other languages, though if we are to provide treatments for conditionals in other languages, we ought to be aware of their salient grammatical features. I believe that much of the trouble with traditional analyses of English conditionals is that they fail to pay proper attention to the grammar of natural language expressions.

My semantics for English conditionals seeks to fulfil three important desiderata for any natural language semantics. The first desideratum is that *a meaning theory for a natural language should be compositional.* According to Lepore and Ludwig, a compositional meaning theory for a language *L* is a formal theory that enables anyone who understands the language in which the theory is stated to understand any sentence in *L* on the basis of understanding the primitive expressions of *L*. A compositional meaning theory is a requirement for any natural language in order for that language to be learnable by finite beings. The second desideratum is that *the meaning theory should preserve "semantic innocence."* Unless we are forced to do so, we should avoid saying that an expression is ambiguous. Natural languages are complex, messy, and full of vague and ambiguous expressions. However, if we can give a semantic treatment of a certain class of natural language expressions that avoids treating those expressions as ambiguous, we ought to do so. There is no need to suppose that natural languages are messier than they really are. The final desideratum is that *the meaning theory should avoid controversial ontological claims.* Unless we are forced to do so, we should not introduce special theoretical entities to explain the semantics of certain expressions. This is Occam's Razor applied to semantics.

In Chapter 2, I attack the traditional *semantic* distinction between "subjunctive"/"counterfactual" and "indicative" conditionals. As the terminology suggests, this distinction is usually made on the grounds of either a conditional's subjunctivity or its counterfactuality. I show that neither means of making the distinction is adequate, and argue that any semantic theory that bases itself on an improperly defined distinction is highly questionable.

The focus of Chapters 3 and 4 is historical. I trace the development of the two most popular sorts of theories of so-called "subjunctive" or "counterfactual" conditionals in the literature. The topic of Chapter 3 is entailment-based theories, such as those of Nelson Goodman (1955) and Roderick Chisholm (1946). Entailment-based theories say something like the following:

> A subjunctive conditional 'If it were the case that P, then it would be the case that Q' is true just in case there is a true law sentence '*L*,' and a true sentence about relevant background conditions '*C*,' such that '$P \wedge L \wedge C$' logically implies '*Q*'.

The precise details of Goodman and Chisholm's theories vary slightly from this, though they both employ the central idea of counterfactual truth requiring some sort of relation of entailment between the antecedent (along with certain other true statements) and the consequent of a conditional. The problem with these theories, as Goodman perceived, is that it is impossible to spell them out without circularity.

Chapter 4 will discuss possible worlds-based theories, such as those of David Lewis (1973) and Robert Stalnaker (1981). I define a *possible worlds* analysis very generally as an analysis that makes use of the notion of a possible world, whatever the expression 'possible world' is used to mean, such that the truth conditions (and perhaps also the semantics) of so-called "subjunctive" or "counterfactual" conditionals are determined by what is or is not the case in those worlds. This sort of analysis clearly has some initial appeal, and it is undoubtedly the most widespread in the literature. However, the use of possible worlds in a theory of conditionals certainly requires some justification. All other things being equal, an analysis that does not refer to such entities seems preferable because of its greater economy of theory.

Of course, possible worlds theorists will deny that all other things are equal. One way of motivating the use of possible worlds is to claim that they capture something about what we actually do when we think about "counterfactual" or "subjunctive" conditionals. We consider one or more hypothetical situations in which a certain counterfactual state of affairs obtains (as described by the antecedent of the conditional under question) and see what else would be the case in those situations. Then we can make a decision about whether to accept the conditional or not. Donald Nute (1980) calls this process "hypothetical deliberation." However, the fact that we undoubtedly go through such a process does not imply that "counterfactual" conditionals are *about* possible worlds. Imagining possible worlds where counterfactual states of affairs are true may be no more than a helpful heuristic device for thinking about these conditionals and deciding which ones we should accept.

My semantics for so-called "subjunctive" or "counterfactual" conditionals is a truth conditional semantics, and in Chapter 5, I describe how such a semantics might be given for a natural language. I follow Lepore and Ludwig's (2000) model for such a semantics, which is based on a suggestion of Donald Davidson that we use Tarski's Convention *T* to help produce a theory of meaning for a natural language. I show how past tense verbs, such as the ones that appear in so-called "subjunctive" or "counterfactual" conditionals, can be treated in such a semantics. The semantics treats tense as an indexical device for referring to time intervals at which events occur or states obtain. Past and future tenses in particular involve what Lepore And Ludwig call *indexically restricted quantifiers*. The indexical element picks out an utterance time relative to which temporally bound states and events are quantificationally located. I compare this approach to the semantics of tense expressions with the more traditional approach that often goes under the name of "tense logic." This approach treats tense expressions as sentential operators rather than quantificational devices. I argue that such an approach is not useful in the analysis of tense expressions in natural languages.

In Chapter 6, I argue for a theory of indicative conditionals. This is important to my cause, since I claim that so-called "subjunctive" or "counterfactual" conditionals are simply tensed indicative conditionals. It is therefore necessary to have a theory of indicative conditionals in place in order to analyze the semantics and conversational role of so-called "subjunctive" or "counterfactual" ones. My analysis of indicatives is based upon the *equivalence thesis.* The equivalence thesis, as defined by Frank Jackson, is the thesis that the material conditional '$(A \supset B)$' is logically equivalent to the indicative conditional '$(A \rightarrow B)$'. I also claim that they are *semantically* equivalent, but Jackson does not. Clearly, on frequent occasions, we would not assert an indicative conditional even when we believe that the equivalent material conditional is true. In particular, we tend not to assert conditionals simply because we know that they have a false antecedent or a true consequent, and it would often be unhelpful or misleading to do so. To explain this phenomenon, I appeal to principles of conversational helpfulness. These principles will tell us that, in most cases, we should not assert a conditional unless we believe that there is some non-truth functional connection between the antecedent and the consequent, e.g., that the proposition expressed by the antecedent entails or nomically requires the proposition expressed by the consequent (Grice refers to this as the *Indirectness Condition*).

In Chapter 7, I elucidate the body of my theory. In place of the traditional terminology, I precisely define a class of conditionals, which I call "backshifted." This class is not distinct from the class of indicative conditionals, since, according to my theory, all conditionals are indicative (strictly speaking, the description 'indicative' is redundant). The class of backshifted conditionals contains all those conditionals that people call "counterfactual" or "subjunctive" as well as a few others. Combining my semantics for tense with my semantics for indicative conditionals, I produce a semantics for backshifted conditionals.

Finally, in Chapter 8, I consider where my theory stands with respect to a number of theorems and argument forms of interest to those who study so-called "counterfactual" or "subjunctive" conditionals. In particular, I consider how my analysis treats a number of argument forms that many philosophers say are invalid for so-called "subjunctive" conditionals. These include Hypothetical Syllogism, Strengthening Antecedents and Contraposition.

CHAPTER 2
THE TRADITIONAL DISTINCTION

If there is a distinction between semantic categories of conditionals, we ought to be able to specify the grounds for that distinction. However, those who employ the traditional distinction between indicative and subjunctive (or counterfactual) conditionals have hardly ever made that distinction clear. Many philosophers and logicians assume the existence of a distinction without attempting to define it. Moreover, those who subscribe to the traditional distinction say little or nothing about how the component expressions of so-called subjunctive (or counterfactual) conditionals contribute to their meanings. The conditionals in question are usually marked by special verb inflections (e.g., the use of 'were' in the antecedent or 'would' in the consequent) or the addition of auxiliary verbs (e.g., to form the past perfect in the antecedent or past shifted future perfect in the consequent). However, the traditionalists remain silent about what role these grammatical features play.

My goal in this chapter is mostly destructive. First, I argue that we can make no interesting philosophical distinction between conditionals on the grounds of either their supposed counterfactuality or subjunctivity. Second, I attempt to diagnose why philosophers have bought into an illusionary distinction.

Counterfactuality

The conditionals in question are often called counterfactuals. Does this imply that they can be distinguished by the counterfactuality of their antecedents? In *Fact, Fiction* *and Forecast*, Nelson Goodman says that "[c]onsidered as truth-functional compounds, all counterfactuals are of course true, since their antecedents are false" (Goodman, 1955, p. 4). However the view the antecedent of every so-called counterfactual conditional is false is not a popular one, and for good reason. Whether or not a conditional's antecedent happens to be false is surely irrelevant to whether or not it belongs in the alleged special category that Goodman and others perceive. It would indeed be odd if a conditional like 'If the match were struck, it would burn' were excluded from the special category if someone merely happened to strike the match in question. What actually happens or not is surely irrelevant to the question of how we determine the meaning or truth conditions of a conditional. The simple action of striking a match could never turn a conditional of one kind into a conditional of a fundamentally different kind whose truth conditions are determined in a fundamentally different way. There are facts that are relevant to the semantics of our sentences, but these are the facts about our practices using words, not about whether a match has been struck or not, or a whether or not a gun has been fired, and so on.

The expression "counterfactual conditional," as it is used in most of the literature, does not mean "conditional with a false antecedent." Those philosophers who use it generally do not want to deny that so-called indicative conditionals can also have false antecedents. For example, the conditional 'If John is from California, then John is American' has a false antecedent, yet I doubt anyone would call it a counterfactual.

None of what I have said denies that the assertion of certain kinds of conditional hints that their antecedents are contrary to fact rather than entailing it. Indeed, this is how Jonathan Bennett sees the traditional distinction:

> [T]he thought is that a 'counterfactual' conditional is one that is worded so as to suggest or presuppose or weakly imply that the antecedent is false: it is infelicitous to utter a 'counterfactual' unless you think its antecedent is false, or are talking to someone who thinks it is false. (Bennett, 1988, p. 509)

We are perfectly able to define a class of conditionals along these lines. However, these are highly dubious grounds on which to mark off a category of conditionals that supposedly have fundamentally different truth conditions. Bennett realizes this:

> A class of conditionals that is marked off by a weak semantic feature – one that concerns felicity of utterance rather than truth, and concerns what is loosely implied rather than what is strictly said or entailed – could also be marked off by something special in the truth conditions, but it would be unwise to assume that it is so. (Bennett, 1988, p. 510)

It is not entirely clear what Bennett means by a "weak semantic feature."[1] However, as I see it, the felicity of a conditional's utterance is a pragmatic consideration, not semantic at all. There is scant reason to believe that such a distinction could mark off a difference in truth conditions.[2] Note also that the utterance of many conditionals traditionally classified as indicative hints at the falsity of their antecedents. For example, if you did not believe it had a false antecedent, it would be highly infelicitous to assert (1).

(1) If the defendant is innocent, then I'm a monkey's uncle.

---

[1] Perhaps he has in mind something like Grice's idea of a conventional implicature, like the sort that is supposed to attach to 'but' to the effect that there is a unexpected contrast of some sort between what is expressed by sentences that flank it.

[2] Even if it were a conventional implicature, it would not mark off a difference in truth conditions, so interpreting "weak semantic feature" in this way would not blunt this consideration.

Here, the deliberate absurdity of the consequent indicates that the speaker thinks that the antecedent is not to be believed. However, nobody would call this conditional a counterfactual or prescribe a possible worlds analysis for it.

In sum, there appears to be no way to capture the intended class of conditionals by appeal to either the actual or implied falsity of their antecedents. The actual truth value of a sentence in the antecedent of a conditional is *independent of* those practices which determine its semantic form. Moreover, many conditionals with false antecedents are clearly *not intended to be in* the favored class. Appeal to the implied falsity of an antecedent fares no better. This looks to be at best a pragmatic, not a semantic feature, and in any case could not make a difference to its truth conditions. And, again, we see that this criterion would subsume conditionals clearly not intended to be in the favored class. The term "counterfactual," then, which is currently the most popular label for the intended class of conditionals, fails to pick out any feature which all and only the conditionals in the intended class share—an astonishing state of affairs, which argues that philosophers who use the term have but a poor grasp on the mechanism by which the conditionals they use manage to express what they intend to express using them.

A Grammatical Distinction?

If we cannot make a distinction on the grounds of counterfactuality, can we make it on grammatical grounds? The use of the expression "subjunctive conditional" suggests that the distinction between the two classes of conditionals is a grammatical one based on the mood of their verbs. However, the grounds for making the distinction this way are dubious. Many philosophers freely talk about a class of "subjunctive" conditionals while simply ignoring all the linguistic evidence against this idea. Unlike many other languages, the subjunctive is barely used at all in Modern English, especially in the spoken language. As Otto Jespersen says in his *Modern English Grammar on Historical Principles*,

> [a]part from certain stereotyped phrases or types of phrases the subjunctive is now little used in spoken language, but tradition has kept certain functions of it alive in the literary language, though in most of these cases it is now possible to use the indicative instead.

The subjunctive still exists in a number of archaic forms. However, most of the examples which philosophers claim to be subjunctive conditionals are more properly treated as cases of what Jespersen calls "imaginative tenses."

> Verbal forms which are primarily used to indicate past time are often used without that temporal import to denote unreality, impossibility, improbability or non-fulfillment. In such cases we speak of imaginative tenses or tenses of the imagination. (Jespersen, 1954-58, Vol. VI, p. 113)

Jespersen gives the following examples, (2)-(4) of "imaginative tenses':

(2) I wish I had enough money to pay you.
(3) If I had money enough, I should pay you.
(4) You speak as if I had money enough.

Utterances of (2)-(4) implicate something about a contrary-to-fact situation. In Old English a subjunctive was required in such cases, but this is not so in the modern language:

> Originally this use was found in the preterit *subjunctive* only, and the unreality was denoted by the mood rather than the tense. But in the course of time the distinction between the forms of the subjunctive and those of the indicative tense came to be blotted out, and now in 99 pct. of cases it is impossible from the form to tell which of the two moods is used.

The only significant form in which the indicative/subjunctive distinction survives in modern English is in the use of singular 'were'. For example,

(5) If I were you, I would buy a smoke detector.

However, Jespersen also points out that even singular 'were' belongs to both the indicative *and* subjunctive mood. This is demonstrated by some examples of Dudman (1988, p. 121). Many philosophers and logicians regard the use of singular 'were' as a definite indication of a subjunctive conditional. Dudman shows that this is demonstrably wrong for singular 'were' is an option in cases where logicians and philosophers recognize no subjunctive. He gives the following examples, (6)-(8):

> (6) He had been as ass to suppose Captain Steerforth a murderer; it was part of the confusion of mind caused by Dr Drake's pills. And if Captain Steerforth were innocent, then he was a potential, a natural ally against his enemies. (Evelyn Waugh, 1957)
>
> (7) If he happened to be in the midtown area at noon, it was a convenient place for lunch, and, if Annabel were playing bridge late, a pleasant one for a solitary evening drink. (Louis Auchinscloss, 1973)
>
> (8) We discussed what was best to be done in our circumstances. Violet put forward a proposal: if Parliament were recalled we would set off for home at once. (Anthony Powell, 1980)

In these cases, singular 'were' is used in conditionals whose antecedents describe epistemic possibilities, not contrary-to-fact situations. For instance, (6) suggests that there is a very good chance that Captain Steerforth is innocent. The conditional in (6) contains a singular 'were', but it does not merely countenance a contrary-to-fact hypothesis in which he is not a murderer. This would lead most philosophers not to recognize a subjunctive in cases like (6), (7) and (8), especially if they believe that the use of the subjunctive in a conditional implies that its antecedent is an imaginary hypothesis.

Consider a conditional like (5). Someone may insist that surely there are subjunctive conditionals and this is one. However, even if this is a subjunctive conditional, the inflection of its verb is not the right way of picking out the intended class of conditionals because other conditionals traditionally placed in the special class do not contain inflected verbs. For example,

(5a) If I had been you, I would have bought a smoke detector.

is traditionally classified as in the special class in question, but the inflection of the verbs does not indicate a subjunctive. Moreover, if it is not subjunctive, (5a) gives us some reason not to take (5) to be subjunctive, because it looks as if it is simply the result of the past shifting of the verbs in protasis and apodosis of (5). This argues for the 'were in the first being a variant of 'was'.

So, in Modern English, past-tensed verbs can be used to indicate (in some sense yet to be determined) counterfactuality, exactly as subjunctive verbs used to do in Old English. This suggests that the mood of the verb (indicative as opposed to subjunctive) is not what philosophers ought to be interested in when attempting to capture the intended class of conditionals.

So far, then, we have seen that there is reason to doubt that the subjunctive survives even in those conditionals thought to be subjunctive, and that, in any case, some conditionals which are intended to fall into the relevant class clearly are not subjunctive. Thus, the label "subjunctive conditional" is just as defective a classifier for the intended category of conditional as “counterfactual conditional.” There is no avoiding the conclusion that no coherent classification of these conditionals has been advanced at all.

To further support this conclusion, we can turn now to consider how we use past tensed verbs for talking about the present or future. The aim of this is to illustrate how we can achieve the effect thought to be achieved by the use of the subjunctive mood indirectly. This will also help to set the stage for my positive account of those conditionals philosophers have sought to classify as “counterfactual” or “subjunctive.”

In considering this, we must be careful to distinguish semantic considerations from merely pragmatic ones. In conversation, we use words in all sorts of ways, and frequently we are not able to interpret a person’s utterance by considering its meaning alone. Here’s an example from Searle (1979). We are sitting at the dinner table, and I say, ‘Can you pass me the salt?’ What I have said clearly has the grammatical form of a question. Now, unless you are incapacitated, you could answer the question truthfully with a simple ‘Yes’. However, if you did so, I would no doubt think that you were either being deliberately facetious or making a joke. I was obviously implicating that I wanted you actually to pass me the salt. Searle gives a plausible account of how the implicature is generated in virtue of the question ‘Can you pass the salt?’ requesting information relevant to a presupposition of a direct request to pass the salt. In the context, where it is clear that you know your interlocutor can pass the salt, it will be clear to him that the information you explicitly ask after cannot be the point of the question. However, whether he can pass the salt certainly is relevant to something else that, in the context, looks to be of interest to you - getting some salt. In particular, it is relevant to a presupposition of requesting that he pass the salt. So, he reasons that you do not really want to know whether he can pass the salt, that being obvious, but what you ask about is relevant to a presupposition of the request that he pass the salt, and it is a good guess at the dinner table that the point of asking about a presupposition of a request that he pass the salt is that you are interested in having the salt passed to you. He is forced to reason this way in order to make sense of what you said and therefore understands that you wanted your utterance to be taken as a (polite) request that he pass the salt (the more direct 'Pass the salt' seeming too blunt). What you did was ask a question, but what you *implicated* was that you'd like the salt passed to you. Here, the conclusion that there is a difference between semantics and pragmatics is difficult to avoid.

Just as we can use a question to make a request, we can use a past tense verb for talking about the present or future. I want to consider what the mechanism is by which we achieve this. I begin with the following suggestion of Jespersen's about how the "imaginative tenses" work:

> [t]he tense which is ordinarily used to express past time here simply removes the idea from the actual present and keeps the action or state denoted by the verb at some distance. (Jespersen, 1954-58, Vol. VI, p. 114)

To simplify matters, I will illustrate this suggestion with a non-conditional example. Suppose my electricity goes out during a hurricane. I say,

(9) I wish I had a flashlight.

Why do I use the past-tense 'had' rather than the present tense 'have' here? Jespersen's idea is that my use of the past tense distances the idea of my having a flashlight from the present situation. The presupposition must be that I would never utter the sentence if I believed that a flashlight were nearby. When we use past tensed verbs to talk about the past, they indicate distance from the present. Jespersen's suggestion is that when we use them to talk about the present, we interpret the idea of "distance" more generally, maybe in terms of spatial distance as well as temporal distance.

However, it's not clear that Jespersen is right about this. There are circumstances in which one could say, 'I wish I had a flashlight', even when one believed one to be close by. For instance, one might say 'I wish I had a flashlight, and if I could reach that top shelf there I'd soon be in possession of one', or 'I wish I had a flashlight, and soon I'll be in possession of one'.

A better idea is that the use of the past tense suggests that my predicament of being without a flashlight derives from events at some time in the past. For example, maybe I stopped by the hardware store yesterday and forgot to pick one up. So I do not have a flashlight now and, presumably, there was some period of time in the past when I did not have one either. To put it somewhat crudely, I say that I wish I *had* one (thereby picking out the past), because if I *had* one then, I might also have one now! 'Wish' can also take 'were' in the complement: 'I wish I were in possession of a flashlight'. If we treat 'were' here as a variant of the past tense 'was' we can employ the same explanation for its force here as well.

Thus, a speaker sometimes uses a past tense verb when it is clear from the context that she intends to say something relevant to the present. It's relevance to the present then is to be understood in terms of our imagining things being as represented in the past, and things unfolding in accordance with what we would normally expect up until the present. This works to explain the force of 'I wish I had a flashlight'. Though the speaker is obviously concerned about the present, she speaks about a wish for something in the past. Her reason for doing this must be that she wants us to imagine how things would have developed form that some imagined past state of affairs. There was some time in the past, at which something could have happened which did not-e.g., the speaker could have remembered to bring a flashlight-in consequence of which, she now does not have a flashlight. By using the past tense 'had' the speaker is asking us to imagine how things would have developed from a past state in which, e.g., she remembered to bring a flashlight that would result in her now having a flashlight. In this way, she expresses her desire to have a flashlight currently, by expressing a wish for something past which if it had obtained would suffice for her to now have a flashlight. In contexts in which such utterances are appropriate, the reasoning is transparent, and without reflection we conclude the primary point of the speaker's utterance is to indicate that she desires that she have a flashlight at the time of utterance.

This explanation of the force of 'had' that takes the speaker to be indicating the relevance of something in the past to the present may be extended to explain the force of 'had' in conditionals as well. Suppose I were to assert 'If I had a million dollars, I would be a rich man'. Why would I utter this? Literally, I am saying that some past time is such that if I had a million dollars at that time, then at some time after that I would be a rich man. You obviously do not understand my utterance as being particularly concerned with the past. The point of it must be to convey something about the present by saying this thing about the past. I am not a millionaire, nor have I ever been, but I clearly did not assert the conditional simply because it is trivially true in virtue of a false antecedent. What then, could my reason have been? You must surely interpret my utterance as expressing something about the connection between having a million dollars and being rich thereafter, regardless of whether it has to do with the past or the present. The effect of a further past shifting, 'If I had had a million dollars, I would have been a rich man', is to have us take the speaker to be intended to talk about the past, but again not about what was so then, since the conditional would not be needed for this, but about connections between hypothetical past events.

It might be objected to this line that 'had' is ambiguous between the subjunctive and simple past of 'have'. For example, one might be tempted to say that 'had' has a different meaning in each of these two sentences:

(10) I had my flashlight.
(9) I wish I had my flashlight.

In the first case, one might claim that 'had' has its regular past-referring meaning while, in the second, it has a special, 'counterfactual' meaning. However, this response is unpromising.

A plausible principle is that, whenever an expression is ambiguous, we should expect it to be systematically ambiguous. In other words, it should be ambiguous in all contexts of the same semantic type in which it occurs (with certain qualifications to be noted). If this is right, whenever a speaker uses an ambiguous expression in a certain grammatical role, we should always be able to interpret his utterance in more than one way. Sometimes, context may disambiguate, as when we have a predicate ambiguous between two senses that apply to different kinds of objects; e.g., 'Mustard is an odd choice', 'Three is an odd number'. A term may also be ambiguous between two different grammatical categories where its grammatical role in the sentence fixes which sense is relevant; e.g., ‘Anne can juggle’, ‘Anne opened a can of soup’. In general, though, we will expect two readings to be available when we have a term ambiguous between two senses which contribute in the same kind of way to a sentence, so that its position in the sentence and the kind of terms it is combined with do not force us to interpret it one way rather than the other. Disambiguation then will depend on context, and both readings will be semantically available. For example, if I were to say, ‘I’m putting my money in the bank’, you would ordinarily interpret me as indicating a plan to deposit my cash in a financial institution. However, in other contexts, you could interpret me as indicating a plan to bury my money by the river (I might be an eccentric who distrusts anyone handling my money, so I bury it in the river bank for safekeeping).

‘Had’ is functioning as a verb in both (9) and (10), and expresses the same relation. If ‘had’ is ambiguous here, the only way of disambiguating it must reside in the inflection. However, any general semantic device attaching to the inflection should be available in any context, but it is not in the case of (10). Someone might insist that its embedding in the complement after ‘wish’ forces one reading, which should not be available elsewhere. However, there are two things to be said against this. First, it is not clear that ‘had’ cannot be interpreted as the simple past in the complement. We can interpret ‘I wish I had a flashlight last night’ as expressing a wish about a past state (though there is some tendency to use the past perfect here: ‘I wish I’d had a flashlight’). Second, we know that the subjunctive in other languages can function outside of embedding in contexts like that in the complement in (9). So whatever its semantic function, that function does not prohibit it from manifesting itself in contexts like (10). Therefore, the proponent of the ambiguity thesis does not have an explanation of why the supposed subjunctive reading is *not* available in a context like (10). This lends weight to the argument that the typical effect is at root pragmatic, though perhaps the form has become typecast, as it were, in something of the same way that 'Can you pass the salt?' or 'Do you have the time?' have, so that it appears to be semantic.

To summarize this section, we considered whether the class of conditionals philosophers have traditionally had in mind in talking of "counterfactual" or "subjunctive conditions" can be marked off by their use of the subjunctive mood in their protasis and apodosis. I argued against this on a number of grounds. First, there is no historical basis for this claim. In modern English, the subjunctive has disappeared from all but archaic uses. The only plausible survival is the use of 'were' in the protasis of conditionals, but even here, since 'were' belongs to the indicative as well, it is not clear that there is any special subjunctive use left. Indeed, I argued the past shifting of a typical 'were' conditional yields a non-subjunctive conditional, which argues that 'were' functions simply as a variant of 'was'. Moreover, I argued at some length that clearly past tensed verbs can be used to indicate something about a time other than the past, either the present or the future, in non-conditional as well as conditional contexts. In non-conditional cases, we can explain the implications of standard uses without supposing the tensed verbs mean anything other than what they do in other contexts. This suggests there is no need to invoke the subjunctive to account for the import of the constructions usually thought to harbor it. I argued also that the thesis that past tense verbs are ambiguous is implausible, because then we ought to expect two readings where there are not two readings of the relevant tensed expressions. The conclusion we should draw is that the relevant class of conditionals is not to be identified by their special use of the subjunctive mood. *Rather, they are simply conditionals that use the past or past perfect in the protasis and the past shifted future or past shifted future perfect in the apodosis.* The considerations advanced in this section strongly suggest that the standard import of such conditionals should be viewed as the result of pragmatic rather than semantic features, the tensed verbs being taken in the standard sense. The challenge then is to say in some detail how the import can be extracted from an independently established semantic account of the tenses, particularly the perfect tenses and the past shift future modal. This task I take up in Chapter 7.

Diagnosis of the Traditional Distinction

One cannot make the traditional distinction between conditionals on the grounds of either their counterfactuality or their subjunctivity, and very few philosophers who use the traditional terminology have attempted to describe the distinction in any other way (I will mention some notable exceptions later). This prompts the question of why philosophers have felt the need to make the traditional distinction at all.

Many traditionalists seem to rely on an instinctive feeling that there are two different kinds of conditional with different truth conditions. However, this instinctive feeling seems to derive from a general failure to distinguish the *meaning* of a conditional statement from other considerations. In particular, there is a failure to distinguish what a conditional means from what we think about when evaluating that conditional. Consider Lewis's example at the beginning of *Counterfactuals*:

> 'If kangaroos had no tails, they would topple over' seems to me to mean something like this: in any possible states of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over. (Lewis, 1973, p. 1)

On most occasions, when we want to assess the truth or falsity of conditionals such as the one about kangaroos, we think about hypothetical situations in which counterfactual states of affairs obtain. On first reading Lewis's example, one is likely to imagine a tailless kangaroo in its familiar surroundings and wonder whether it would be able to able to keep its balance. This seems to motivate Lewis's claim that, when we use conditionals like this, we are literally talking about hypothetical situations (or possible worlds[3]). Other conditionals, the so-called indicative ones, do not seem to imply anything about hypothetical situations. For instance, an utterance of the conditional 'If Joey is a kangaroo, Joey is a marsupial' does not invoke any thoughts about hypothetical kangaroos. Therefore, the reasoning goes, we must have two different kinds of conditionals with fundamentally different semantics.

However, the assumption that the idea of a hypothetical situation tells us anything about the semantics or truth conditions of conditionals like Lewis's seems to be based on an unwarranted inference. It might seem a small step from admitting that these conditionals make us think about hypothetical situations to the conclusion that they are *about* hypothetical situations (or maybe that their truth conditions concern hypothetical situations). However, this is not the case. Rather, thinking about hypothetical situations could merely be a heuristic device to help evaluate conditionals and decide which we should accept and which we should reject. For instance, if I were to say, 'The bear is a lumbering animal', you might imagine a bear moving around and ask whether you would describe it as 'lumbering'. This might help you decide how to react to what I said, but my statement was clearly not about imaginary or hypothetical bears. Similar considerations could apply in the kangaroo example. We imagine a world with tailless kangaroos to help us determine how we should respond to an utterance of the conditional 'If kangaroos had no tails, they would topple over', but this does not entail that the conditional is *about* hypothetical situations. Of course, I have not yet established the conclusion that the conditional is *not* about hypothetical situations. I have merely challenged the legitimacy of the inference from the indisputable claim that certain kinds of conditional cause us to think about hypothetical situations to the contentious conclusion that they are actually *about* these situations. This unwarranted inference has driven many philosophers toward the traditional distinction.

Another misguided motivation behind the traditional distinction derives from a failure to pay proper attention to the constituent parts of conditional sentences and, in particular, the tense and mood of their verbs. This has caused many to misidentify the antecedents and consequents of conditional statements.

Most philosophers and logicians analyze conditionals into three components: an antecedent, a consequent and a conditional operator. They perceive that the conditional operator in certain conditional sentences does not function like a regular indicative, so they infer that a different operator must be involved, i.e., a special "subjunctive" or "counterfactual" one. This creation of a special counterfactual operator has caused logicians to perceive further apparent differences between conditionals. For example, it appears that hypothetical syllogism, contraposition and strengthening antecedents are invalid argument forms for conditionals with a counterfactual operator (these argument

---

[3] I will use the expressions 'hypothetical situation' and 'possible world' as synonymous.

forms are discussed in more detail in Chapters 3 and 8). However, the traditional argument for discerning a counterfactual operator is not convincing.

Many traditionalists subscribe to Lewis's argument for the conclusion that there are two distinct kinds of conditional requiring different analyses. We are asked to compare the following two sentences (originally introduced into discussion by Ernest Adams):

(11) If Oswald didn't kill Kennedy, somebody else did.
(12) If Oswald hadn't killed Kennedy, somebody else would have.

According to conventional thinking, (12) is a "counterfactual" or "subjunctive" while (11) is an ordinary indicative. Most people would agree that (11) is clearly true, while (12) is contentious. It makes sense to assert (11) while denying (12). The problem is to explain this difference in reaction to the two sentences.

Many philosophers and logicians, including Lewis, claim that the antecedent of both (11) and (12) is

(a) Oswald didn't kill Kennedy

and the consequent of both is

(b) Somebody else did [kill Kennedy].

This leads quickly to the conclusion that (11) and (12) are different kinds of conditional construction. If their antecedents and consequents are the same, little else could explain our difference in reaction. Yet, what motivates the assumption that the antecedent of (12) is (a) and the consequent is (b)? It seems a curious thing to suppose, because the inflection of the verbs that appear in (a) and (b) is different from those in the original sentence. One perceived difference between (11) and (12) is that (11) can be viewed as two complete sentences, i.e., (a) and (b), joined together by the 'if . . . then . . .' construction whereas (12) does not contain two such sentences.

However, this is surely incorrect. Given the right sort of context,

(c) 'Oswald hadn't killed Kennedy'

and

(d) 'Somebody else would have [killed Kennedy],'

are exactly sorts of constructions that can stand alone as sentences outside of conditional constructions. Adverbials usually accompany the verb constructions in these sorts of sentences. For instance:

> (13) It was October 1963. Oswald hadn't yet killed Kennedy.
> (14) Oswald awoke on the morning of November 22nd with a feeling of apprehension. By the end of the day, he would have killed Kennedy.

In (13), the tense of the verb in 'Oswald hadn't yet killed Kennedy' is the past perfect. An utterance of (13) implies that there is a time in the past at which 'Oswald has not killed Kennedy' was true. From the context, we can see that the time in question was October 1963.

In (14), the verb construction 'would have' is a relativized future modal, the so-called future perfect in the past. The sentence 'I will have finished lunch in twenty minutes' is of a similar form, but not shifted into the past. Its utterance implies that the sentence 'I have finished lunch' will be true in twenty minutes, relative to the time of the utterance. In (14), the relativization is to a time in the past made clear by the surrounding context. In this case, it is the morning of November 22nd. This relativization explains why it is difficult to think of cases in which someone would say 'Oswald would have killed Kennedy' all by itself. The verb construction 'would have' necessarily involves a relativization to another time in the past, identified from a surrounding conversational context. When 'would' or 'would have' appears in the consequent of a conditional, I suggest that the relativization is to a time picked out by the antecedent.

Since (c) and (d) are sorts of constructions that *can* stand alone outside of conditional constructions, the reason for not identifying them as the antecedent and consequent of (12) must be that, whatever semantic function the verbs in (c) and (d) perform in (13) and (14), it has nothing to do with the role they play in (12). Perhaps one motivation for this thought is that that conditionals with similar verb constructions to (12) are sometimes used to talk about hypothetical events in the future, not just the past. For instance, suppose two of your friends have just called off a marriage scheduled for next month. Relieved by the news, you say:

(15) If they had married, they would have been unhappy.

One might think that the verb construction 'had married' could not be the past perfect in this case because the sentence is not used to talk about the past. Moreover, it is not difficult to find other examples of the simple past being used to talk about the present or future or the present tense being used to talk about the future.[4] For example, Dudman points out that each of the following conditionals has a future interpretation:

> (16) If she misses the last bus, the Countess will walk home.
> (17) If she missed the last bus, the Countess would walk home.
> (18) If she had missed the last bus, the Countess would have walked home.
> (uttered just after learning that the Countess was assassinated on the way to the bus stop) (Dudman, 1989, p. 595)

On one interpretation, (17) says something about the Countess's habitual behavior in the past, as in 'Whenever the Countess missed the bus she would walk home'. On its future interpretation, (17) is similar in import to (16). Note also that (16) has an interpretation according to which it talks about habitual behavior in the present, as in 'Whenever the Countess misses the bus she walks home'. Conditional (18) has a future interpretation if it is uttered just after learning that the Countess was assassinated on her way to the bus stop.

Let us call conditionals that exhibit this phenomenon 'backshifted'. In backshifted conditionals, we use verbs for talking about times later than their tenses would suggest in their ordinary, non-backshifted contexts. The phenomenon of backshifting seems to have led many to the traditional distinction. They see the past perfect being used to talk about the future and think to themselves 'that couldn't be the past perfect, it must be a special subjunctive or counterfactual construction!'

Interestingly, the phenomenon of backshifting leads Dudman to make a distinction in a different way. Correctly identifying their use of tense, he argues that conditionals like (16), (17) and (18) ought to be marked off from the likes of

---

[4] Strictly speaking, English does not have a future tense. So there is a very real sense in which we *must* use other tenses for talking about the future.

(19) If she missed the last bus, the Countess walked home.

Dudman argues that (19) is a compound sentence, comprising two separate sentences joined together by the 'if' construction. However, he argues that (16), (17) and (18) are simple sentences. "[T]he sentence," he says, "divides into a subject 'The Countess', and a predicate containing everything else" (Dudman, 19xx, p. ?). The clause whose first word is 'if' becomes part of the predicate. Thus, on its future interpretation, (16) is more perspicuously represented as (20),

(20) The Countess will-if-she-misses-the-last-bus walk home.

Traditionalists would categorize both (16) and (19) as indicatives, placing them on the other side of the distinction from (17) and (18). Dudman places (16), (17) and (18) on the same side of the distinction, contrasting them with (19). I find it hard to see how Dudman motivates this move. Presumably, he believes that (17), for example, is not a compound of 'she missed the last bus' and 'the countess would walk home' because he does not see how the past-tensed verbs in the two compound sentences could perform the same semantic function as they do in (17). On the other hand, an utterance of (19), standardly interpreted, says something about the past, as do its compound parts 'she [the Countess] missed the last bus' and 'the Countess walked home'. Therefore, the same apparent difficulty does not occur.

However, both Dudman and the traditionalists are wrong. From a semantic perspective, past-tensed verbs pick out times in the past, but this does not restrict their use in conversation to talking about the past. I have already shown how this suggestion might cash out in some non-conditional cases, and how this might be extended to conditional cases. In order to show how this sort of proposal can work in detail in the case of so-called counterfactual conditionals, I will first need to develop a semantics for tense expressions and a thorough analysis of how pragmatic considerations function in the utterance of conditionals.

To summarize this section, we have seen that there are two major causes behind philosophers' dubious allegiance to the traditional distinction. The first cause is a failure to distinguish semantic considerations from pragmatic ones in our use of conditional sentences. The fact that the utterance of a conditional causes us to think about certain imaginary situations, as it does with a "counterfactual," does not imply that it is *about* such situations. The second cause is a failure to perceive the antecedents and consequents of conditional sentences properly. Lack of sensitivity to the tense expressions that occur in these sentences causes philosophers to reconstruct their antecedents and consequents in such a way that leads to the conclusion that some conditionals have special "counterfactual" operators.

Summary

The traditional distinction between the two supposed classes of conditionals, one traditionally called indicative, and the other counterfactual or subjunctive, can neither be made on grammatical grounds, based on the indicative/subjunctive distinction, nor on the grounds of counterfactuality, implied or actual, of their antecedents. No useful semantic distinction can be made on traditional lines at all. Syntactically, what sets apart the relevant class of conditionals is the combination of verb forms in the protasis and apodosis of the conditionals. The relevant class of conditionals comprises those that have the past or past perfect in the protasis (counting 'were' as a variant of 'was') and the past shifted future or future perfect, respectively, in the apodosis. The failures of the tradition to adequately characterize the intended class of conditionals is directly linked to their failure to take seriously (a) the tense structure of the protasis and apodosis in the relevant conditionals, because they (b) have not adequately taken into account the possibility that the features of the use of these conditionals they are interested in are the result of the interaction of their semantics with pragmatic factors. And we have seen the that special features of the use of such conditionals which have motivated their treatment as a special class semantically are quite plausibly construed instead as pragmatic in origin instead. We have discovered there are both conditional and non-conditional expressions in which the past tense is used to do different things. As well as its ordinary use in talking about the past, the past tense can be used to talk about hypothetical events in the present and future. It can also be used to suggest counterfactuality. This leads us to the conclusion that any distinction to be made between kinds of conditionals is more likely to be pragmatic than semantic. Therefore the most constructive way to approach a semantics of the intended class of "special" conditionals is to come first to an understanding of the meaning and use of *tense* in the relevant conditional statements. Before I take up this task directly, however, I will examine in more detail the two dominant traditional positions on the semantics of so-called counterfactual conditionals, in the next two chapters, against the backdrop of the discussion of this one.

CHAPTER 3
LINGUISTIC ANALYSES

In this chapter, I will be concerned with so-called linguistic theories of "counterfactual" or "subjunctive" conditionals. All such theories trace back to Nelson Goodman's seminal treatment of counterfactuals in *Fact, Fiction and Forecast*. The salient feature of all such theories is that they see the truth, or, more broadly, acceptability, of counterfactual conditionals as grounded in a certain kind of argument that is to be or can be constructed from the supposed antecedent and consequent of these conditionals, with the antecedent figuring as one of the premises, and the consequent as the conclusion. I will consider first Goodman's introduction of the problem of counterfactual conditionals, which presupposes, falsely, as I have claimed, that such conditionals should be thought of as employing a special sentential connective, and then the difficulties that arise for his own attempt at a solution to the problem he poses, namely, the entailment-based theory. Goodman's theory runs aground on the problem of specifying non-circular cotenability conditions (this will be explained below). Certain more sophisticated descendants of Goodman's theory, cast in the form of a possible worlds account, I postpone discussion of until the next chapter. After discussing Goodman style theories, I turn to a variant of the idea that rejects the view that counterfactual conditionals are truth evaluable. I discuss in particular J. L. Mackie's well-known account. I argue that the main feature of such theories, the denial of the truth evaluability of counterfactual conditionals, is counterintuitive, inadequately supported, and cannot accommodate our embedding of such conditionals in truth functional compounds. I also suggest that it is at least doubtful that they make any advance on the problem of cotenability that plagued Goodman's theory.

In *Fact, Fiction and Forecast*, Nelson Goodman describes the Problem of Counterfactual Conditionals as the problem of defining "the circumstances under which a given counterfactual holds while the opposing conditional with the contrary consequent fails to hold." (Goodman, 1955, p. 4) In saying this, Goodman influenced the way many philosophers approach the analysis of so-called counterfactuals. Consider, for instance, Goodman's own example about a piece of butter that was eaten yesterday:

> (1) If that piece of butter had been heated to 150°F, it would have melted. (Goodman, 1955, p. 4)

Goodman thinks the problem is to specify when this conditional holds and the following conditional does not:

> (2) If that piece of butter had been heated to 150°F, it would not have melted. (Goodman, 1955, p. 4)

Any thought that we can treat (1) and (2) as simple material conditionals by giving them a truth functional analysis is quickly dispelled. Goodman says that, "[c]onsidered as truth-functional compounds, all counterfactuals are of course true, since their antecedents are false" (Goodman, 1955, p. 4). So, if we analyze counterfactuals in this way, both of the above are true. For similar reasons,

> (3) If that piece of butter had been heated to 150°F, it would have grown into an elephant.

is also true. Goodman clearly thinks this is a *reductio ad absurdum* of a truth-functional analysis.

Goodman never explicitly says what he thinks the antecedents of (1) and (2) are. However, from what he says elsewhere, he appears to think they are both 'that piece of butter is melted'. Most other writers would identify them as 'that piece of butter was melted'. Since neither the phrase 'that piece of butter is melted' nor 'that piece of butter was melted' appears in (1) or (2), there is clearly some sort of reconstruction going on here. In fact, almost every theory of so-called counterfactual conditionals ever conceived depends on this sort of reconstruction of antecedents. As I explained in Chapter 2, I disapprove of this type of maneuver. However, during this chapter and the next, I will simply assume that the antecedent of (1) and (2) is 'that piece of butter was melted'. I have two reasons for this. The first is simply convenience, to avoid clumsiness of expression. Whenever, I use the term 'antecedent' in Chapters 2 and 3 with reference to a so-called counterfactual or subjunctive conditional, it would more appropriately be replaced with 'what most authors consider to be the antecedent'. The second reason is that I want to criticize these theories from an internal perspective. By an 'internal' criticism, I mean a criticism that accepts the assumptions of the theory in question and discovers problems with how that theory is then worked out. Clearly, if we do not assume that conditionals have antecedents and consequents in the way that Goodman supposes, his analysis fails before it begins.

During this chapter and the next, I will assume that the traditional distinction between indicative and subjunctive is correct. Again, this is partly for convenience, and partly because the theories discussed presuppose the distinction, and I want to highlight problems that occur independently of the terminological issue. The reader might ask, if there really is no distinction to accept, then what exactly am I assuming here? I certainly agree that we are unable to precisely delineate a class of 'subjunctive' or 'counterfactual' conditionals. However, we have several examples of the sort of conditionals that people who use the distinction call 'subjunctive' or 'counterfactual', so we have plenty of cases to test their theories against. For example:

> If Oswald had not killed Kennedy, somebody else would have. (Ernest Adams)
> If the vase were dropped on the floor, it would break. (Roderick Chisholm)
> If you had had the brakes fixed, there would not have been a collision. (J.L. Mackie)

I will therefore evaluate these theories according to whether they offer a satisfactory analysis of sentences such as the above. When I use the expression 'counterfactual conditional', it can be considered shorthand for 'so-called counterfactual conditional' (similarly for 'subjunctive conditional' and 'so-called subjunctive conditional').

Goodman presumes too much when he says that the Problem of Counterfactuals concerns the circumstances under which conditionals like (1) and (2) hold true. Obviously, we react differently to (1) and (2). An utterance of sentence (2) provokes disagreement in a way in which (1) does not. Goodman assumes that the explanation for this difference in reaction is that (1) is true and (2) is false. Yet, there could be other explanations for our reaction to (2). In Chapter 7, I argue we react unfavorably to conditionals like (2) on pragmatic grounds (of course, when *Fact, Fiction and Forecast* was written, Goodman did not have access to Grice's groundbreaking work on the pragmatics of conversation). To acknowledge the possibility of alternative explanations, and not to beg any questions, it would be better to describe the Problem of

Counterfactuals as the problem of explaining the difference in reaction that we have to utterances of (1) and (2), not the problem of explaining the conditions under which the two conditionals hold true.

Having said this, it is still rather obvious that we cannot treat counterfactual conditionals as plain and simple (unquantified) material conditionals. On this kind of analysis, all sorts of strange-sounding conditionals turn out to be true. This objection alone would not prove fatal to this simple kind of theory if it were able to provide some grounds for thinking that, though these conditionals are really true, we have some good reason for not asserting them, or even mistakenly believing them to be false. However, such a simple analysis does not give us adequate resources to explain why these conditionals are deemed unassertable or otherwise unacceptable on any sort of pragmatic grounds.

Another proposal quickly ruled out in most of the literature is that one can treat counterfactuals as simple, universally quantified material conditionals. Such a proposal would say, for instance, that a counterfactual like

(4) If kangaroos had no tails, they would topple over.

should be treated as

(5) For any x, if x is a kangaroo, then x topples over.

The reason why most find this unacceptable is that there appears to be a difference between accidental universal generalizations and law-governed counterfactuals. Many writers claim that many counterfactuals are true because they are supported by causal laws, i.e., because there is some law-based connection between the antecedent and the consequent of a counterfactual conditional. Yet, a universally quantified material conditional can be true by accident. For example, given that

> (6) For any x, if x is a former American President, then x is a man

is true, the proposal under consideration finds that

> (7) If Hilary Clinton were a former American President, then Hilary Clinton would be a man.

is also true. Most people would think anyone uttering (7) would thereby say something false , and even if (7) is not really false, it is certainly not something most people would typically assert. Once more, the fact that it takes (7) to be true need not prove a fatal objection to the analysis if some reason can be found that explains our usual reaction to it, but the problem again is that the analysis does not provide us with adequate resources to give such an explanation. With the aid of additional resources and somewhat more subtlety, I believe that an analysis of counterfactuals that treats them as quantified material conditionals can be made plausible. However, there is more to do before we will be in a position to advance such an analysis. I return to it in Chapter 7.

Early Entailment-Based Theories

Nelson Goodman and Roderick Chisholm were interested in the analysis of counterfactual conditionals largely because of their application in philosophy of science.

At the beginning of *Fact, Fiction and Forecast*, Goodman says that if we cannot properly analyze this kind of conditional, we cannot lay claim to any adequate philosophy of science. Both Goodman and Chisholm note that we express many scientific laws using disposition terms. These include not only those terms ending in words like 'uble', 'ible' and 'able' (e.g., 'soluble', 'malleable', 'combustible') but also, says Goodman, "almost every objective predicate, such as 'is red'." (Goodman, 1955, p. 3) It is clear enough that the use of disposition terms generally licenses beliefs about contrary-to-fact situations. As Chisholm notes, to say that a thing is fragile ('fragile' is a disposition term) is to say that, under certain conditions, which may or may not be realized, it would break (Chisholm, 1946, p. 291). In addition, Chisholm emphasizes the use of these conditionals in historical claims, e.g., "If we had followed a different policy toward Germany in the 1920's, the Second World War would never have occurred" (Chisholm, 1946, p. 290). Most importantly, what we think about these kinds of claims might affect our decisions in the future. Whatever else we believe about science and history, there would be little purpose to these disciplines if they did not allow us to say things about what would happen should certain results obtain regardless of whether they actually do. Indeed, our ability to express beliefs about such things is important in almost every aspect of our lives. Thus, Goodman and Chisholm acknowledge that a precise analysis of these conditionals is a problem of central importance, and their efforts represent the first serious attempts to produce such an analysis.

Goodman and Chisholm were right to emphasize that our understanding of disposition terms involves beliefs about how objects to which they apply behave in circumstances actual and non-actual, holding fixed natural law. Undoubtedly, there is also a relation between sentences containing disposition terms and counterfactual conditionals. However, the nature of this relation is not immediately clear. It could be that sentences containing disposition terms are synonymous with certain counterfactual sentences. Alternatively, the relation might be one of entailment (and the entailment could go in either direction) or logical equivalence. The relation might be weaker still-it might simply be that an utterance of a statement containing disposition terms licenses the hearer to believe certain contrary-to-fact claims without the statement's actually entailing them. According to the view that I advance, it will be an assertion to the effect that some object is, e.g., fragile, will entail something that will be standardly implicated by the use of an appropriate counterfactual conditional.

Goodman claims that "[a] counterfactual is true if a certain connection obtains between the antecedent and the consequent" (Goodman, 1955, pp. 7-8). The antecedent is supposed to play the role of a ground for the consequent. The strongest sort of ground would be secured if the antecedent entailed the consequent. However, Goodman notes, "the consequent seldom follows from the antecedent by logic alone" (Goodman, 1955, p. 8) - it needs to be conjoined with some other sentences. These sentences express what Goodman calls the *relevant conditions* for the conditional in question. Specifying exactly what these sentences are while avoiding circularity is perhaps the biggest obstacle facing an entailment-based theory.

There are a number of accounts of counterfactuals that have much in common with Goodman's. These include those of Roderick Chisholm (1946) and Wilfred Sellars (1975). These accounts are sometimes called "linguistic" or "metalinguistic" accounts. Lewis (1973) calls them "metalinguistic" because "they hold that linguistic entities –

arguments and their premises – enter prominently into the truth conditions or assertability conditions of counterfactuals" (Lewis, 1973, p. 66). John Pollock (1976) defines a *linguistic* analysis of subjunctive conditionals as any analysis that says that a counterfactual conditional '$P > Q$', where we represent the supposed counterfactual connective with '>', is true just in case there is a true law sentence 'L', and a true sentence about relevant background conditions 'C', such that '$P \wedge L \wedge C$' logically implies 'Q'. 'C' expresses the *relevant conditions* for the conditional in question. The exact details of some of these theories differ in subtle ways from this, but all employ the central idea that a true counterfactual conditional requires some sort of entailment relation between its antecedent (along with certain other true statements) and its consequent.

The most obvious problem with these analyses is in determining exactly what belongs in the set of relevant conditions for a given counterfactual conditional (that is, what we put in the place of our schematic letter 'C' above). It quickly becomes obvious that we must put some restrictions on what can be included in this set. It cannot simply be the conjunction of all true sentences, otherwise the theory finds many counterfactual conditionals true which appear false. For instance, if 'If it were the case that P, then it would be the case that Q' is a counterfactual conditional with a false antecedent, then 'not-P' is a true circumstance and will therefore be included in 'C'. Since absolutely anything follows from the conjunction of P and 'not-P', the theory will find any counterfactual conditional with a false antecedent to be true. It seems, therefore, that we must at least require that C be consistent with the antecedent of the conditional in question. In addition, without restriction on C, it turns out that whenever the material conditional '$P \supset Q$' is true, '$P > Q$' is true also. This is because if '$P \supset Q$' is true then it must be included in *C*, in which case *C* conjoined with *P* logically entails *Q*. (As with the other simple theories mentioned above, that this theory finds some conditionals true that first appear false might not constitute a fatal objection if some explanation can be given for this circumstance, but it does not appear that an entailment-based analysis without any restriction on relevant conditions could provide any such explanation.)

Differences in entailment-based analyses are most likely to be differences in what sorts of requirements they put on relevant conditions. One sort of requirement that seems unavoidable is what Goodman called a *cotenability* requirement. Consider a counterfactual conditional '$P > Q$'. Goodman argued that any sentence included in the set of its relevant conditions should not be false if *P were* true. In other words, if *R* is included in the set of relevant conditions for this conditional, then *R* must be *cotenable* with *P, where R* is *cotenable* with *P* just in case '$-(R > -P)$' is true. It is not enough simply to require that all the statements in the set of relevant conditions be logically consistent (or even physically compossible) with the truth of *P*. Suppose, for instance, that Lou takes the bus to work. However, suppose that he hates busses and only uses them because he cannot afford to buy a car, and that there is no other available means of transportation. In such a situation, we would not assert the following counterfactual:

> (10) If Lou were to win a million dollars in the lottery, he would take the bus to work.

It is true that Lou takes the bus to work, so, if we have no cotenability requirement, we can put this in the set of relevant conditions for the counterfactual. Then, the antecedent, together with the set of relevant conditions, entails the consequent and the counterfactual is found true – an unwanted result. This is not the sort of example Goodman would have envisioned, since he was only interested in the use of counterfactuals in the sciences, but it illustrates the problem very clearly. Mackie reconstructs Goodman's own example illustrating this problem as follows (Mackie, 1962, p. 76). Suppose we have the following five beliefs:

(a) All dry matches located in an oxygen-containing medium light when struck.
(b) M is a dry match.
(c) M is located in an oxygen-containing medium.
(d) M has not been struck.
(e) M has not lit.

Suppose also that we are asked to evaluate the following counterfactual claim:

(11) If match *M* had been struck, it would have lit.

If we take the antecedent of (11) together with (a), (b), and (c) as relevant conditions we can directly infer the falsity of (e). Thus, (11) is found true. However, without a cotenability requirement, we can also take the antecedent together with (a), (c) and (e) and directly infer the falsity of (b). Thus the following is also found true:

(12) If match *M* had been struck, it would not have been dry.

However, for Goodman, (12) must surely be false. Using a similar trick, one could also infer (13) and (14) from the antecedent together with the above five claims:

(13) If match M had been struck, it would not have been the case that all dry matches located in an oxygen-containing medium light when struck. (using (b), (c) and (e))
(14) If match M had been struck, it would not have been located in an oxygen-containing medium. (using (a), (b) and (e))

Notice that, in (12), (13) and (14), (e) is used in the inference to the consequent. The problem is that (e), though compatible with the antecedent, is not something we would ordinarily say is true. In other words, it is not cotenable with the antecedent and must be ruled out of the set of relevant conditions on those grounds.

The need for a cotenability requirement in an entailment-based theory is obvious. If we use relevant conditions to determine what would follow from the truth of *P*, we need this set to include only those sentences that would be true if *P* were true. Allowing it to include sentences that would be false if *P* were true leads, intuitively, to the wrong results. However, if an entailment-based theory puts a cotenability requirement on set of relevant conditions for a counterfactual conditional, there is an immediate risk of circularity, because cotenability itself is defined using a counterfactual conditional.

Goodman could not see an effective way of solving this problem. He suggests revising his theory of counterfactuals so that it becomes a sort of foundational theory. The idea would be to admit first those counterfactuals whose truth can be inferred from their antecedents alone, that is, without the need for relevant conditions (their antecedents would have to directly entail their consequents). It is not clear exactly what these counterfactuals would be like, though it seems likely that they would have to be sentences describing molecular physical properties. One would then use these counterfactuals as criteria for the cotenability of relevant conditions of other counterfactuals. However, Goodman is far from satisfied with this suggestion because he wants a "step by step method" for evaluating counterfactuals in the sciences. He says that if one is forced to examine all the foundational counterfactuals in order to determine the truth of a simple counterfactual conditional such as 'If the match had been struck it would have lighted', the process will quickly become horrendously complex (Goodman, 1955, p. 17).

Whether or not we require a "step by step" method for deciding the truth of a counterfactual, Goodman is right that this sort of strategy is not promising. It is radically divorced from the way we think about counterfactuals in everyday situations. It is quite clear that, when we attempt to evaluate the truth of a counterfactual like 'If the match had been struck it would have lighted', we do not give any thought to counterfactuals which are true based on their antecedents alone. Such a process would just take too long. Though the objection is not necessarily fatal, a theory that fails to explain how we think about counterfactuals in everyday situations is not promising.

Later Entailment-Based Theories

Despite their problems, entailment-based theories did not die out with Chisholm and Goodman, though later incarnations are somewhat more complex, usually employing model structures or systems of possible worlds. In most instances, if one thinks about possible worlds as sets of sentences or functions from sentences onto truth values (rather than metaphysical entities), a possible worlds-based theory reduces to an entailment-based theory. Theories such as John Pollock's (1976) and N. A. Blue's (1981) can be seen as entailment-based theories when regarded in this sort of way. Roughly speaking, these theories say that a counterfactual is true just in case its consequent is true in certain possible worlds in which its antecedent is true together with some subset of laws and some (cotenable) set of sentences. If we think of the set of statements that are true in these possible worlds similarly to sets of relevant conditions, the similarities with entailment-based theories are not hard to see. A counterfactual is true just in case, in certain possible worlds, picked out by the laws and sentences true in them, its consequent is true whenever its antecedent is true. This is much the same thing as saying that the antecedent, together with the laws and other sentences must entail its consequent. The circularity problem then reduces to the problem of picking out the right set of possible worlds so that everything that is true in those worlds is jointly cotenable. Pollock attempts to solve the problem by considering just those possible worlds in which certain maximally consistent sets of "basic strong subjunctive generalizations" hold true, where these generalizations are defined inductively. However, theories like Pollock's are too complex to discuss here. I consider them in greater detail in Chapter 3 under the category of possible worlds-based theories.

"No Truth Value" Theories

There is another group of theories of counterfactual conditionals which is often placed in the "metalinguistic" category. For want of a better name, I call these "no truth value" theories, since they claim that counterfactual conditionals are neither true nor false. J.L. Mackie's (1962) provides the most well known example of this kind.

Mackie describes counterfactual conditionals as incomplete and condensed arguments. Though philosophers contest whether truth and falsity are properties of sentences or propositions, they are never properties of arguments. So, if counterfactual conditionals are really arguments, it is simply a category mistake to describe them as either true or false.

Prima facie, Mackie says, a conditional like (15),

(15) If you had had the brakes fixed, there would not have been a collision,

describes a situation while hinting that it is imaginary (Mackie does not impose any requirement that a conditional's antecedent be false in order for it to be considered a counterfactual). He claims that the description in this conditional suggests that a situation in which the brakes have been fixed and a collision was avoided has not come about. However, he believes there is another element to the above conditional, namely that it suggests that the absence of a collision somehow "follows from" or "goes with" the brakes being fixed (though not in such a way that the thing suggested is truth evaluable). In other words, it suggests that there is some argument that could be given, with 'you had the brakes fixed' as a premise and 'there is no collision' as a conclusion. The conditional, however, does not give the inference in full. Other premises and intermediate steps in the argument are suppressed. Thus, Mackie claims, counterfactual conditionals are *incomplete* arguments. In the case of the conditional about the brakes, he suggests that the argument might go as in (16):

> (16) Suppose that you had the brakes fixed. Then when the other car turns across your path you press the brake pedal. So your car stops quickly. So there is no collision. (Mackie, 1962, p. 68)

The argument is still incomplete. To complete it, says Mackie, we have to add further premises describing the situation in further detail. For instance, we must describe the relevant laws connecting the pressing of the brake pedal with the stopping of the car (when the brakes are fixed properly), and those connecting the stopping of the car, in certain circumstances, with the absence of collision. Yet, Mackie does not believe that, in order to understand a counterfactual, one must be able to complete its argument. He claims that we can understand a counterfactual without knowing its grounds.

So, according to Mackie, what is it that we actually do when we assert a counterfactual conditional? When I give an argument, I assert the premises and the conclusion, typically on the understanding that I believe they are true. Yet, when I assert a conditional like 'If you had the breaks fixed, there would not have been a collision', I neither assert 'You had the breaks fixed' nor 'There is no collision'. It is apparently for this reason that Mackie denies that, when one asserts this conditional like this, one is actually giving an argument, or even saying that there is an argument from the premise 'You had the brakes fixed' together with some other true premises to the conclusion 'There is no collision'.

> To advance a counterfactual is not to assert any proposition, even one about an argument; it is rather to run through a condensation of an argument. We can do this without being able to specify explicitly the other premises and intermediate steps on which we are relying; we jump from the supposition to the conclusion in the light of knowledge and beliefs that we need not and commonly do not make explicit. Similarly, we can understand a counterfactual advanced by someone else without either being able to complete his argument or one of our own. In either case the argument is not argued, but rather entertained: not all the premises are asserted, nor is the conclusion. One premise is merely supposed, and the argument from it to the conclusion lies within the scope of this supposition. (Mackie, 1962, pp. 68-9)

Moreover, says Mackie, there might be some indeterminacy in the way the condensed argument is filled out. So, it appears, there is not necessarily a single argument to go with each single counterfactual conditional. In effect, on Mackie's view, when we put forward a counterfactual conditional, such as 'If you had had the brakes fixed, there would have been no collision', we are sketching an argument without filling it in, as if we had said:

(17) You had the brakes fixed at t1, etc. Therefore, there is no collision at t2

As I have already noted, for Mackie, counterfactual conditionals cannot be said to be true or false. Instead, they are said to be "sustained by the premise or premises that, with the antecedent as a further premise, entail the consequent of the conditional" (Mackie, 1962, p. 69). So, the important distinction is between those conditionals *sustained* by certain premises and those that are not. Yet, if the distinction between sustained and unsustained conditionals is to be useful, we must surely put some restriction on which premises we are allowed to use in the inference from the antecedent to the consequent. If a conditional could be sustained by adding any premises of one's choosing (whether true or false) to the antecedent, every counterfactual would be sustainable. Therefore, says Mackie, the premises must be drawn from our other knowledge and beliefs. Also, given the belief-contravening nature of most counterfactual conditionals, when searching for the premises that will sustain them, we must not include everything we believe to be true, but only what is consistent with the truth of the belief-contravening suppositions that usually form their antecedents.

Mackie thinks his analysis of counterfactual conditionals leaves us with two problems. First, we must explain under what circumstances we are prepared to advance a conditional, and, second, we must say why we are justified in doing so. It is not enough to require simply that true premises sustain the conditional. The problem, for Mackie, is that many accidentally true generalizations fail to sustain counterfactuals. Suppose we have discovered, by checking with each individual, that every person in the room understands English. Thus, the generalization 'Every person in the room understands

English' is true. However, the truth of this generalization does not seem to license an assertion of the counterfactual 'If Mr. Krushchev were in the room, then he would understand English' (the example is Mackie's own). Mackie seeks to explain why the generalization does not sustain the counterfactual with the following principle:

> A generalization sustains a counterfactual if our reason for adhering to it is not undermined when we add the supposition. (Mackie, 1962, p. 72)

According to Mackie, the counterfactual, 'If Mr. Krushchev were in the room, then he would understood English', is a condensed argument which takes us from the hypothetical premise, 'Krushchev is in the room' to the conclusion, 'Krushchev understands English'. Mackie claims that the reason why the generalization 'Every person in the room understands English' does not sustain the counterfactual is that our reason for believing the generalization to be true is undermined by the supposition that 'Krushchev is in the room' is true. Given that Krushchev does not understand English, and the only reason we believe that generalization to be true is that we have checked that each individual understands English, our reason for believing that the generalization is true disappears.

Ultimately, for Mackie, the problem of explaining why a causal law sustains counterfactuals reduces to the traditional problem of induction. This he freely admits.

> [T]o explain why a causal law sustains counterfactuals is to explain how there can be evidence for a universal proposition which is not impaired either (i) by a supposition that there is an instance of the subject term which has not been excluded in the evidence, or (ii) by a supposition which adds further (contrary-to-fact) instances of the subject term. (Mackie, 1962, p. 73)

Mackie does not attempt to provide a solution to the above problem, but it would not be fair to criticize his theory on these grounds. However the problem is to be solved, that solution can be plugged in to Mackie's theory.

On the face of it, it is really rather puzzling what Mackie thinks we do when we assert counterfactuals. Suppose I assert a counterfactual on no grounds whatsoever, such as the following:

(18) If France were to become a monarchy, Australia would become a republic.

We will assume there is no hidden connection between the politics of France and Australia that would make (18) a legitimate assertion. According to Mackie, by asserting this counterfactual, I have "entertained" an argument. If it were a good argument, with the right sort of premises, the counterfactual would be sustained and I could assert it legitimately. However (18) is obviously not a legitimate assertion (given what we suppose, it would not make sense as a move in most easily imaginable conversational contexts). Have I therefore entertained a bad argument? It seems to me that I have not entertained any argument at all. I have just asserted something at random which I do not even believe to be true. However, what I asserted certainly seems to have some content, and there might even be a purpose to my asserting it, such as to mislead someone. In light of this, Mackie's theory is not very plausible. I certainly might be in error when I deny that I'm entertaining an argument, but if I am, I at least require an explanation of why I would believe such a thing incorrectly, and no such explanation is forthcoming from Mackie.

Mackie claims as a virtue of his theory, however, that it copes with many of problems that plague entailment-based theories such as Goodman's. Consider again the problem posed by the example of the match. We have the following five propositions which we believe:

(a) All dry matches located in an oxygen-containing medium light when struck.
(b) M is a dry match.
(c) M is located in an oxygen-containing medium.
(d) M has not been struck.
(e) M has not lit.

The problem is to explain, without circularity, why we endorsing these allows us to assert (19), but not (20), (21) or (22).

(19) If match M had been struck, it would have lit.
(20) If match M had been struck, it would not have been dry.
(21) If match M had been struck, it would not have been the case that all dry matches located in an oxygen-containing medium light when struck.
(22) If match M had been struck, it would not have been located in an oxygen-containing medium.

Mackie's thought is that the denial of (d) in the antecedents of all these conditionals rules out our grounds for believing (e). Therefore, since they rely on (e), (20), (21) and (22) are not well-sustained counterfactuals. (19) does not rely on (e), so it is assertable while (20), (21) and (22) are not.

However, it is not clear that Mackie succeeds in avoiding the difficulty, rather than hiding it. Why is (d) grounds for believing (e)? Surely this is not something that is primitive. It is not enough to say that there is a causal law connecting match strikings with their lighting, because that is not the only way matches can be lit. Rather, we need to know that the match *would* light in the circumstances only if it *were* struck. Given this, then knowing that the match is not struck is reason to believe it is not lit. But this is to appeal to a conditional of the same general type as that which we want to explain. It is not clear, then, that Mackie has avoided the problem of circularity that exercised Goodman, rather than displaced it.

There are, in any case, additional worries about Mackie's proposal that attach to other aspects of it. The most commonly made objection to theories of counterfactuals that deny that counterfactuals have truth values is that they are unable to give a proper analysis of compounds with embedded counterfactuals. If counterfactual conditionals do not have truth-values, it is difficult to see how they ought to be treated when they occur in truth-functional compounds such as conjunctions, disjunctions and material conditionals. For slightly different reasons, it is difficult to see how we ought to treat counterfactuals embedded within counterfactuals.[1] There seem to be three possible ways one could counter such an objection (these possibilities map the logical space of possible responses and are not necessarily what anyone would argue):

(a) Deny that we use compounds with embedded counterfactuals and that they are therefore ungrammatical or otherwise illegitimate.
(b) Deny that the compounds in question are really truth functional.
(c) Deny that the compounds in question are truth functional when they contain embedded counterfactuals.

Strategy (c) is just *ad hoc,* and strategy (b) fares no better. Whatever grounds anyone might have for denying that conditionals are truth functional, to claim that no disjunctions or conjunctions operate truth-functionally is just implausible. There might be grounds for denying that *some* disjunctions and conjunctions operate truth functionally,[2] but there are surely no grounds for denying that any of them are. The only remotely plausible strategy is (a), but after a little reflection, this quickly begins to look unpromising too.

---

[1] Apparently similar problems occur for certain theories of indicative conditionals which accept the standard truth functional analysis and explain the strangeness of certain conditionals, like 'If Elvis is alive, he's living in Belgium' on pragmatic grounds while maintaining that they are true on truth functional grounds. These theories include those of Grice, Jackson and myself. I deal with these problems in Chapter 6.

Strategy (a) is not convincing because it is quite easy to find examples of compounds with embedded counterfactuals. For instance:

> (23) If Oswald had not shot Kennedy, then Kennedy would have lived, and if Kennedy had lived, he would have completed his term in office.

The following also sound perfectly normal:

> (24) If Oswald had not shot Kennedy, somebody else would have, and everybody knows it.
> (25) If Oswald had not shot Kennedy, somebody else would have, or I'm very much mistaken

Sentence (26) is an example of a counterfactual embedded in a an indicative conditional:

> (26) If I'm not very much mistaken, then if Oswald had not shot Kennedy, somebody else would have.

Sentences (27) and (28) contain counterfactuals embedded in counterfactuals:

---

[2] For example, in *Relevant Logic*, Stephen Read draws a distinction between *intensional* and *extensional* disjunctions and conjunctions. The extensional ones are to be treated truth-functionally while the intensional ones are not. The distinction is based on what grounds we have for thinking the compound expression to be true. Imagine, for instance, that a fox is chasing a rabbit. He comes to a fork in the road, and tries to reason which way the rabbit went. He says to himself "the rabbit either took the left fork or the right fork." If his grounds for saying this are that he actually saw the rabbit take the left fork, it is an extensional disjunction; it's truth has been inferred from the truth of one of its disjuncts. If he simply reasons that the left fork and the right fork are the only two roads the rabbit could have taken though he did not see which road it actually took, it is an intensional disjunction; it's truth has *not* been inferred from the truth of one of its disjuncts. According to Read, this distinction has particularly important consequences when considering the validity of disjunctive syllogism as a mode of argument.

> (27) If McCain wins [were to win] in South Carolina, then if he were [were he] to return to using negative ads in the Arizona and Michigan primaries, he would certainly win the Republican nomination.
> (28) If I were to become [became] a politician, then I'd win an election only if I were to forget all my training, and my morals.

These examples show that we clearly do sometimes use compounds with embedded counterfactuals and that strategy (a) therefore fails.

It is quite easy to find examples of conditionals with conditionals as antecedents, but is perhaps less clear whether we ever use conditionals with conditionals as antecedents. One might be tempted to argue that we do not use conditionals with conditionals as antecedents. Consider, for example, (29):

> (29) If if we can go then the van is fixed, then we'd better take it to the shop.

This certainly sounds awkward. However, (29) is standardly treated as equivalent to (30), which seems unproblematic.

> (30) If we can go only if the van is fixed, then we'd better take it to the shop.

Why do we prefer (30) to (29)? One obvious suggestion is that the two consecutive *if*s in (29) are clumsy and are difficult to pronounce. Rearranging the conditional with the 'only if' equivalence avoids this problem.

From these various examples, we can see that we embed conditionals in all sorts of different constructions, including other conditionals. Thus, all three of the strategies mentioned fail, from which we can conclude that Mackie's "no truth value" theory offers us no adequate response to the worry that it cannot give a proper treatment of embedded conditionals.

Summary

Where does this leave us? Linguistic theories come in two basic varieties, the entailment- based theories and "no truth value" theories. In summarizing the results of this chapter, I divide the problems that arise for these theories into problems with their procedure, which I will call 'Methodological Problems', and problems with their analyses, which I will call 'Analytical Problems'.

Let us consider the methodological problems first. Both sorts of theories have a strike against them at the outset because they ignore the grammar of the conditionals they analyze. They take themselves to have antecedents and consequents, which are not licensed by the surface grammar, but rather read into the conditionals on the basis of intuitions about what speakers are trying to convey by using them. This is a dangerous procedure at best, particularly in the light of what we have learned from Grice about the pervasiveness of systematic conversational implicature.

A connected methodological problem is the fact that neither sort of theory raises or explores the possibility that the intuitions we have about what's intended by utterances of counterfactual conditionals may be intuitions about what is standardly implicated by uses of them given a semantic analysis which takes seriously the apparent tense of the protases and apodoses.

The explanation of these methodological failings no doubt is in part connected with the fact that these theories were first formulated prior to Grice's work on conversational implicature being fully assimilated into work in the philosophy of language. Grice's William James lectures were given at Harvard in 1968, and only first appeared in print in a short article in 1975. Goodman's work, which set the tone for subsequent discussions, was published in 1949. The history of discussion of "counterfactual" conditionals might have been much different if Grice's work had been available earlier, or serious work on counterfactuals begun after Grice's work. Once a view about the shape of a problem has attainted currency in the philosophical literature, it can acquire the aspect of received wisdom, and remain unquestioned even when later work shows there to have been methodological flaws in the formulation of the problem space originally. This is what I believe to have happened in the case of discussions of "counterfactual" conditionals.

Turning to the analytical problems, we have seen that the main difficulty for entailment-based theories is the problem of cotenability. The problem is that in saying what additional statements we should add to the antecedent in evaluating whether it together with those statements entails the consequent, we must require that they be statements that would be true if the antecedent is: but this makes use itself of a counterfactual conditional, and hence makes the analysis circular. The best repair to this as we have seen would make the evaluation of counterfactuals completely opaque to us in ordinary circumstances, and so deprives this analysis of any power to explain our ordinary practices.

The main difficulty for the "no truth value" theory is that it flies in the face of our practices: we do in fact treat assertions of counterfactual conditionals as being true and false. If someone asserts: 'If Oswald hadn't killed Kennedy, Lyndon Johnson would have', most people would suppose that he has said something false. Moreover, I certainly have no sense that in asserting an arbitrary counterfactual conditional, 'if the sky were red, I'd be a rich man', that I have put forward an argument of some sort, though what I say certainly does not seem to be a piece of nonsense. In addition, as noted, we do use counterfactual conditionals in molecular sentences (sentences formed using the standard logical connectives), which commits us, on the standard analysis of the connectives, to saying that they are truth evaluable. In addition to this, as we saw, it is not clear that the no truth value theories succeed in responding successfully to the cotenability problem.

I will not say that these problems show decisively that no theory along these lines could be correct. But they provide a powerful reason to look elsewhere. In the end, I will argue that we can obtain an adequate theory that avoids the problems, both methodological and analytical, of these linguistic theories.

Though these theories fail, they do, undoubtedly, uncover something about how we think about the conditionals in question. When we assert conditionals, we generally require that there be some sort of connection between the antecedent and consequent (I do not like the expressions ‘antecedent’ and ‘consequent’ and will replace them with different ones later, but they will do for now). However, this requirement need not be a requirement for their truth, since truth and assertability are not the same thing. There are many true sentences that we would never assert because they are unhelpful or misleading, and some false sentences that have a useful role in conversation. The connection in question might be a law-like or stronger connection between what is expressed in each clause (consider, e.g., ‘if today were Tuesday, tomorrow would be Wednesday’), or it might be that we assert them because we want to draw attention to some argument we have in mind or indicate the possibility of constructing some argument involving the “antecedent” as premise and “consequent” as conclusion. Surely, Goodman and Mackie have captured something about the assertability conditions of counterfactuals, though there is good reason to doubt that their analyses provide the semantics or truth conditions of counterfactuals. We can, as I will argue, both provide a more satisfying account of their literal truth conditions which observes the methodological principles overlooked in the formulation of the problem by Goodman, and which accommodates what is right about the approach of the linguistic theories.

CHAPTER 4
POSSIBLE WORLDS ANALYSES

As noted in Chapter 1, I define a *possible worlds* analysis of so-called counterfactual conditionals as an analysis that makes use of the notion of a possible world, whatever the expression 'possible world' is used to mean, and gives the truth conditions (and perhaps also the semantics) of those conditionals in terms of what is or is not the case in those worlds. This chapter is historical, but also critical. I advanced my fundamental criticism of possible worlds theories in Chapter 2, where I argued that they make an unwarranted assumption that there are two distinct semantic categories of conditional. My criticisms in this chapter are strictly internal (and therefore of somewhat less significance), since they accept this unwarranted assumption. If I did not accept this assumption here, I would not be able to talk about these theories at all, since they presuppose it from the outset.

I assume here that the main business of possible worlds theories is to provide an analysis of the truth conditions of "counterfactual" conditionals. However, it is important to note at the outset that the theories we will consider do not distinguish between the truth conditions of the relevant conditionals and the conditions under which it is appropriate to assert them. My own theory of so-called counterfactual conditionals exploits the distinction between the truth conditions of these conditionals and their assertability conditionals, so that the truth conditions may be different from what one might initially have expected. In my case, this is acceptable because I explain why I believe the truth conditions are the way they are and show how we can derive the standard interpretations of utterances of these conditionals from pragmatic considerations (see Chapter 7). An internal criticism of "counterfactual" theories will not exploit this possibility because these theories do not exploit this distinction. Therefore, I will be considering whether these theories account properly for the "truth conditions" of the relevant class on conditionals on their own standards, standards which, in my view, embody a mistake about what is being tracked. That is, I do not take the standard import of such conditionals to be an accurate guide to what their truth conditions are interpreted literally relative to the occasion of utterance. However, in the following, I am going to pretend, for the sake of mounting internal criticisms of possible worlds analyses, that the standard import of such conditionals is a guide to their truth conditions. I then adopt for the rest of this chapter an official fiction to that effect when I use the expressions 'truth condition' or 'truth conditions'. The reader is hereby warned, however, that in my view this *is* a fiction, and that properly viewed these criticisms of the possible worlds analyses are really criticisms of their correctly tracking the standard conversational import of utterances of the relevant conditionals.

Possible worlds theories of counterfactual conditionals[1] became popular in the late sixties, just a few years after Saul Kripke developed his semantics for modal logics.

---

[1] I argued in Chapter 2 that I do not believe there is a properly defined class of 'counterfactual' or 'subjunctive' conditionals. In this chapter, as in the previous one, I merely assume that this terminology is legitimate for the sake of argument and the sake of convenience. I also assume that we can identify the antecedents and consequents of this class of conditionals in the traditional manner.

Robert Stalnaker and David Lewis thought that the notion of a possible world would be useful in the analysis of conditionals. Stalnaker and Thomason's theories were theories of conditionals in general, both indicative and counterfactual, while Lewis's theory was a theory of counterfactuals alone. Whatever Stalnaker and Thomason's motivations may have been, in this chapter, I am solely interested in whether these theories provide an adequate analysis of counterfactuals.

The basic idea behind possible worlds theories of counterfactuals is well-illustrated in the opening sentences of Lewis's *Counterfactuals*, which I have quoted above:

> 'If kangaroos had no tails, they would topple over' seems to me to mean something like this: in any possible states of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over. (Lewis, 1973, p. 1)

Lewis would not have begun his book with such a statement unless he thought it would strike a chord with his readers, and it undoubtedly does. On most occasions, when we want to assess the truth or falsity of these conditionals, we think about hypothetical situations in which counterfactual states of affairs obtain. This motivates Lewis's contention that when we use these conditionals, we are literally talking about possible worlds and that whether a counterfactual conditional is true or false depends on what is true or false in certain possible worlds.

My main line of my external criticism attacks the assumption that the idea of a possible world tells us anything about the *semantics* of so-called counterfactual conditionals. In Chapter 2, I tried to show where this assumption comes from. I conceded that, on first reading Lewis's example, one is likely to imagine a tailless kangaroo in its familiar surroundings and wonder whether it would be able to able to keep its balance. However, I also argued that the fact that we go through such a reasoning process does not imply that counterfactual conditionals are *about* possible worlds or even that possible worlds have anything to do with the truth conditions of counterfactuals conditionals. Instead, possible worlds might simply be, as far as the evidence goes, a heuristic device for thinking about these conditionals and deciding which we should assert and which we should deny in standard conversational contexts.

One rather weak motivation for a possible worlds-based account is that many other accounts of counterfactual or subjunctive conditionals can be transposed into such accounts and find all the same conditionals to be true and false. Lewis, for instance, suggests how this might be done with Goodman's entailment-based account. Of course, supporters of entailment-based accounts are unlikely to want to transpose them into possible worlds accounts because they are unlikely to be happy with an ontology of possible worlds. However, one might believe that the possibility of such transposition provides some support for possible worlds theories, particularly if one construes possible worlds talk in an ontologically uncontroversial way, e.g., as talk about sets of sentences or propositions rather than about literally worlds filled with possiblia. One can imagine a possible worlds theorist saying to an entailment theorist that his theory is *at least* as good in one respect, because it can mirror the truth values that the entailment theory gives for counterfactual conditionals (though a possible worlds theory obviously differs concerning what is being said about the truth conditions).[2] However, if one is skeptical about both possible worlds and entailment based theories, the possibility of transposition provides little comfort. I contend that both kinds of theory capture something about the way we think about conditionals, but that they both react to pragmatic rather than semantic phenomena.

Material Conditionals and Strict Conditionals

Before we look at possible worlds theories themselves, it will be useful to consider some theories of counterfactual conditionals that Lewis and others find inadequate. These theories include those that treat counterfactual conditionals as material conditionals and as strict conditionals. Theories like Lewis's can be seen as a reaction to the alleged inadequacies of these theories.

Possible worlds theorists claim that the truth of a counterfactual conditional depends not only on the actual truth values of its (supposed) antecedent and consequent (however the antecedent and consequent are identified) but also on their truth values in certain other possible worlds. They therefore deny that we can treat counterfactual conditionals as ordinary material conditionals. Moreover, they also deny that we can treat counterfactual conditionals as *strict* conditionals. Strict conditionals can be given a possible worlds analysis according to which they are true just in case their consequences are true in *any* possible world in which their antecedents are true. Thus, we have the following biconditional:

---

[2] Obviously, by the same token, an entailment theorist could say that his theory mirrors the truth values that possible worlds theories give for counterfactual conditionals. Thus, such an argument is unable to prove that

(1) $(\phi > \psi) \longleftrightarrow \Box(\phi \rightarrow \psi)$[3]

When we define counterfactual conditionals according to this biconditional, our logic will contain all substitution instances of the following theses (since they are provable in even the weakest of modal logics):

(2) Transitivity: $((\phi > \psi) \wedge (\Gamma > \phi)) \rightarrow (\Gamma > \psi)$
(3) Contraposition: $(\phi > -\psi) \rightarrow (\psi > -\phi)$
(4) Strengthening antecedents: $(\phi > \psi) \rightarrow ((\phi \wedge \Gamma) > \psi)$

However, as far as English subjunctive conditionals go, possible worlds theorists argues that there are counterexamples to these theorems. As a counterexample to Transitivity, consider the following three sentences:

> (5) If Carter had not lost the election in 1980, Reagan would not have been president in 1981.
> (6) If Carter had died in 1979, he would not have lost the election in 1980.
> (7) If Carter had died in 1979, Reagan would not have been President in 1981.

Possible worlds theorists argue that utterances of (5) and (6) are true, but it is not at all clear that an utterance of (7) is true. They take the following sets of conditionals to be counterexamples to contraposition and Strengthening Antecedents respectively:

> (8) If it were to rain heavily at noon, the farmer would not irrigate his field at noon.
> (9) If the farmer were to irrigate his field at noon, it would not rain heavily at

---

one kind of theory is better than the other.

[3] '□' represents the sentential operator 'It is necessary that'.

noon.

(10) If Otto had come, it would have been a lively party.
(11) But if both Otto and Anna had come it would not have been a lively party.

These allegedly invalid arguments are used to motivate the claim that counterfactual conditionals are not strict conditionals. The fact that we have unfavorable reactions to these arguments might not prove a fatal objection to the analysis of counterfactual conditionals as strict conditionals if some reason can be found that explains these reactions. However, we have the same problem here that we have with the analyses of counterfactual conditionals as simple material conditionals and as quantified material conditionals that I discussed in Chapter 3. The analysis simply does not provide us with adequate resources to provide such an explanation.

Similarity-Based Theories

Lewis, Stalnaker, and many other possible worlds theorists, employ a "similarity" relation between possible worlds in their analyses of counterfactual conditionals. In this section, I describe these theories and highlight some internal problems with them. Most of these internal problems are related to an inability to adequately specify the notion of similarity that these theories employ.

Though Lewis claims that counterfactual conditionals are significantly weaker than strict conditionals, he wishes to retain his intuition that a counterfactual conditional is true just in case its consequent is true at every member of some *restricted set* of possible worlds where its antecedent is true. Thus, one might still view the semantics of counterfactual conditionals as a development out of Kripkean semantics for modal logics.

Such semantics employ the notions of possible worlds and an *accessibility* relation between them.[4] A number of possible worlds accounts of counterfactual conditionals, including Stalnaker's and Lewis's, also employ a notion of overall similarity between possible worlds.

Stalnaker claims that a counterfactual conditional is true just in case it's consequent is true at the world at which its antecedent is true that *differs minimally* from the actual world:

> This is how to evaluate a conditional: First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is then true. (Stalnaker, 1968, p. 3)

Stalnaker thinks that we can think of possible worlds epistemically as hypothetical situations. Thus, the basic idea is that we modify our beliefs as little as possible in order to accommodate the antecedent of the conditional under question. We can then see whether the consequent would be true if the antecedent were true as in the imagined situation. Somewhat similarly, Lewis says that

> [A] counterfactual 'If it were that A, then it would be that C' is (non-vacuously) true if and only if some (accessible) world where both A and C are true is more similar to our actual world, overall, than is any world where A is true but C is false. (Lewis, 1979, p. 56)

In the model structures for Lewis's semantics, overall similarity is represented by a closeness relation between possible worlds. The closer one world is to another, the more similar those worlds are.

---

[4] The accessibility relation is an artifact of this style of semantics. I do not claim to know how to interpret it.

Stalnaker defines a selection function *f* that takes a proposition and a possible world as arguments and a possible world as its value. In informal terms, when given the antecedent of a conditional *P* and a world *i*, the function is intended to pick out the world most similar to *i* at which *P* is true. Stalnaker gives truth and falsity conditions for counterfactual conditionals (relative to possible worlds) as follows.

$P > Q$ is true at i if Q is true at $f(P, i)$
$P > Q$ is false at i if Q is false at $f(P, i)$

One peculiarity of Stalnaker's account is that it is premised upon the assumption that there is always a unique closest possible world at which the antecedent of a conditional is true (if the antecedent is logically possible). There can never be a tie for the most similar world. In Stalnaker's semantics, this is ensured by the fact that the selection function picks out just one world, not a set of worlds. The assumption that there must be a uniquely most similar world is commonly called the *Uniqueness Assumption.* This controversial assumption is the main difference between Stalnaker's account and Lewis's. In addition to rejecting the Uniqueness Assumption, Lewis also rejects the somewhat weaker *Limit Assumption* which says that, for any sentence *P*, which describes a logically possible state of affairs, there is always at least one world at which *P* is true which is as much like i as any other world at which P is true.

Lewis's rejection of the Uniqueness Assumption is closely connected to his denial of the theorem of *Conditional Excluded Middle* (CEM). In fact, Lewis's rejection of both the Limit and Uniqueness Assumptions in his favored semantics for counterfactual conditionals is what ensures that CEM is not a theorem of Lewis's conditional logic **VC**:

CEM: $(P > Q) \vee (P > \sim Q)$

Lewis argues that there may be no unique *P*-world[5] that is closer to *i* than any other *P*-world. In support of this, he asks us to consider a straight line printed in a book and to imagine that this line were a little longer than it actually is. No matter what length we choose for the line, there will be a shorter length that is still greater than the length of the actual line. Worlds that differ from the actual world only in the length of this line will be more like the actual world the closer the length of the line is to its actual length. However, none of these worlds is the closest at which the line is longer. This can be offered as evidence against both the Uniqueness Assumption and the weaker Limit Assumption.

Cases such as the following, due to Pollock, are often presented as counterexamples to excluded middle. Pollock asks us to imagine that there are two kinds of gasoline, long mileage gasoline (LMG) and short-mileage gasoline (SMG). A tank of LMG enables Pollock to drive to Boston, but a tank of SMG does not. Consider the following conditional, which remains unspecific about the type of gasoline used:

(12) If the car had a full tank, he would be able to drive to Boston, or if the car

---

[5] That is, a world at which P is true.

had a full tank, he would not be able to drive to Boston

Since there are two ways that each antecedent might be true (the car might have a full tank of either type of gasoline), both disjuncts might be false. Therefore, it is argued, the disjunction as a whole could be false, proving CEM not to be a valid theorem. However, this counterexample is not convincing. There are many things that could fill a gas tank other than gasoline, like sand or water. The expression 'had a full tank' in (12) therefore underspecifies a state of affairs relative to which we are interested in whether certain outcomes obtain. The example's reliance on the underspecification in the expression 'full tank' does not make for a convincing case. If we are allowed to change precisely what we mean by 'full tank' in mid-sentence, the sentence 'If the car has a full tank, he can drive to Boston, and if the car has a full tank, he cannot drive to Boston' ceases to be a contradiction.

Using an expression's underspecification to create a counterexample is rather like relying on an expression's ambiguity. For instance, nobody would say that '$p \supset p$' is not valid because 'John is on the bank $\supset$ John is on the bank' has a false reading when 'bank' is interpreted differently in its antecedent and consequent. The question of whether CEM is a valid theorem therefore remains undecided. I will return to the issue of CEM in Chapter 8, after I have advanced my own theory of so-called counterfactual conditionals.

Many different selection functions will satisfy the requirements that both Stalnaker and Lewis place on them, and not all of them are acceptable. Stalnaker considers the possibility that additional formal constraints might be put upon *f*, but he does not think that these constraints will ever be enough to determine a unique function. Stalnaker claims that there might be any number of different but equally acceptable selection functions and that semantic considerations alone will not be enough to pick out a unique one. This refusal to select a unique function is connected to Stalnaker's view that subjunctive conditionals are *pragmatically ambiguous*. For Stalnaker, one cannot resolve this ambiguity by more closely analyzing the meaning of the conditional. It must be resolved by pragmatic considerations relating to the context in which it was uttered. The point seems to be that pragmatically ambiguous sentences have a kind of indexical or demonstrative element and that their truth conditions are fixed only relative to context, in the way in which a sentence containing a demonstrative has its truth conditions fixed only relative to a speaker's demonstrative intention and successful demonstration. Lewis (1973) and Nute (1980) also claim that counterfactual conditionals are pragmatically ambiguous.

I certainly agree that with Stalnaker and Lewis that the context of utterance can make a difference to how a conditional is interpreted. However, I disagree with the claim that this sort of “ambiguity” entails that the conditional itself has two different meanings. (A difference in selection function along the lines that Lewis and Stalnaker suggest implies a difference in meaning, because it implies that the truth or falsity of the conditional in question is decided by the truth or falsity of its antecedent and consequent in a different set of possible worlds.) Grice shows us clearly enough how utterances of the same expression can have a different import in different conversational contexts without resorting to the claim that the expression has a multiplicity of meanings.

Pollock resists Stalnaker's claim that there is any thoroughgoing ambiguity in subjunctive conditionals beyond what one would expect from natural language expressions in general. Pollock cites a number of examples of pairs of conditionals that have been traditionally thought to illustrate the pragmatic ambiguity in question. Perhaps the most commonly mentioned of these conditionals are counteridenticals (conditionals whose antecedents are false identity statements). Consider the following two conditionals:

(13) If Richard Nixon were Golda Meir, then he would be a woman.
(14) If Golda Meir were Richard Nixon, then she would be a man.

Both appear to have logically equivalent antecedents and impossible consequents (though whether the consequents are impossible depends on whether the gender of the pronoun is taken to contribution to the truth conditions of sentences in which they are used). Since the two conditionals cannot felicitously be asserted at the same time, one may be led to believe that some sort of deep-rooted ambiguity prevents us from saying whether either of these conditionals is true without knowing the context of its utterance. Similarly, if I were to point at a rock, I might assert either one (though not both) of the following two different conditionals:

(15) If that *were* gold, it would be malleable.
(16) If *that* were gold, some gold things would not be malleable.

Though both of these conditionals have the same antecedent, you could not felicitously assert them both at the same time. However, each of them might seem like reasonable things to say in certain contexts. This might suggest that there is some form of thoroughgoing ambiguity involved here, so that we can only know when either conditional is true given the specific context in which it is uttered.

According to Pollock, the way to test whether ambiguities in subjunctive conditionals are really of the thoroughgoing sort that Stalnaker supposes is to see whether we cannot regiment or precisify our language to resolve the ambiguity. He thinks that Stalnaker's claims can be resisted if the ambiguity rests upon our simply not being precise enough about what we mean when we use counterfactual conditionals. Pollock's strategy is to employ paraphrases that precisify conditionals and resolve their ambiguities. If Stalnaker were right, such paraphrasing would not be possible, because they claim that it can only be resolved by the conversational context. Pollock maintains that (13) and (14) can be paraphrased as follows:

> (13a) If Richard Nixon had the non-relational properties presently possessed by Golda Meir, then he would be a woman.
> (14a) If Golda Meir had the non-relational properties presently possessed by Richard Nixon, then she would be a man.

Pollock argues that if such a paraphrase is acceptable, it turns out that (13) and (14) really are not counteridenticals and that their antecedents really are not logically equivalent. Therefore, both (13) and (14) can be true at the same time (though only if the consequents can be true). Pollock also claims that the ambiguity in (15) and (16) can be resolved if we pay attention to where the emphasis is placed in the sentence (as italicized). He believes that the difference in emphasis causes a difference in meaning and that it is therefore reasonable to paraphrase (16) as follows:

(16a) If some gold things were like that (i.e., had the properties that has), then some gold things would not be malleable.

Having paraphrased (16) in this way, it is no longer incompatible with (15) and there is nothing to stand in the way of saying that these conditionals have precise meanings. There might still be difficulties about their being true though, on the possible worlds account, since gold might be thought to be essentially malleable, since it is essentially the element with atomic number 79, and arguably what this means is not separable from thinking of it in terms of the correct physical theory governing it, from which it would follow that it was malleable.

Pollock is right that an utterance of (13) has a different import than an utterance of (14), and that this seems to be the case, no matter what the context. For instance, an utterance of (13b) would be infelicitous, regardless of context:

(13b) If Richard Nixon were Golda Meir, then she [Golda Meir] would be a man.

Even if the antecedent of this conditional is an identity statement, swapping the position of the two names makes a difference to the import of an utterance of (13). It seems that, in conditionals like (13), we regard the consequent as saying something about the person named first in the antecedent, so an utterance of (13) causes us to imagine Richard Nixon taking on the properties of Golda Meir, rather than vice versa. This is why an utterance of (13b) would seem odd. However, we can admit this without also admitting that (13) and (14) have different truth conditions, as (13b) and (14b) do, and we have good reason to avoid this claim, since it is far from obvious that these (13) and (14) are synonymous with (13a) and (14a). A more conservative explanation is that, to avoid ambiguity, we adopt a convention about how the consequent is to be understood in a counteridentical conditional relative to the individuals named in the antecedent such that we understand the consequent as saying something about the individual that is first named. However, we do not have to suppose that this convention has anything to do with the truth conditions of the conditional.

There are two main problems with Lewis's theory and *any* theory based upon an overall comparative similarity relation between possible worlds. The first is that it gets the wrong results in a number of cases. Pollock shows how Lewis's argument against the Limit Assumption can be slightly modified to provide counterexamples to similarity-based theories. Suppose that the length of the line in the book is 1". There are then worlds in which the length of the line is between 1" and 1 + x" (for any x greater than zero) which are more similar to the actual world than are those worlds in which the line is 1 + x" (assuming that the length of the line is infinitely divisible). All other things being equal, a world where the line is closer in length to 1" will be more similar to the actual world than one in which the length of the line is a little longer. This assumes that similarity in the length of lines has some bearing on overall comparative similarity between worlds, but I doubt that Lewis would deny this. Since the similarity analysis says that a counterfactual conditional is true just in case it is true at the world(s) most similar to the actual world at which its antecedent is true, it follows, based on the analysis, that the conditional 'If the line were more than an inch long, it would not be 1 + x inches long [for any x greater than 0]' is true. This seems plainly false. Pollock points out that this argument in fact requires a principle not valid in Lewis's semantics, which he calls the *Generalized Consequence Principle*:

> GCP: If S is a set of sentences, and for each $s \in S$ '$P > S$' is true, and S entails R, then '$(P > R)$' is true.

However, Pollock rightly claims that Lewis cannot take much solace from this. Even if GCP is not valid in Lewis's semantics it is a principle that intuitively *ought* to be valid, particularly since the following *Finite Consequence Principle is* valid in Lewis's semantics:

> FCP: If S is a *finite* set of sentences, and for each $s \in S$ '$P > S$' is true, and S entails R, then '$(P > R)$' is true.

Surely, it ought to be possible to generalize from the truth of FCP to the truth of GCP. At least Lewis does not offer any reason why we cannot.

Yet, even without the assumption of the Generalized Consequence principle and the assumption that the line is infinitely divisible, the similarity-based analysis yields bizarre results. Based on considerations similar to those just mentioned, given that I have six coins in my pocket, the following conditional is found to be true:

> (17) If I had over six coins in my pocket, I would have exactly seven coins in my pocket.

It would surely be infelicitous for me to utter such a conditional. If I were to put more coins in my pocket (for instance, after winning at a slot machine), nothing would prevent me from putting more than one more in there, as (17) seems to suggest. However, unless our intuitions are irrelevant, any world in which I have just one extra coin in my pocket is, *ceteris paribus* more similar to the actual world than any world in which I have, say, ten more. It is difficult to see how Lewis can avoid having this conditional come out true on his analysis unless he is prepared to distort his notion of overall similarity out of any intuitively recognizable shape.

The lesson we must draw from these cases is that comparative overall similarity is not the relation involved in determining which sorts of hypothetical situations we must look at in determining whether a counterfactual conditional is true or false. These problems with Lewis's account motivate Pollock (1976) to argue that what is involved in deciding whether a counterfactual conditional is true is not the notion of similarity between worlds but the notion of what Pollock calls *minimal change*. According to this notion of change, a change that leads to a world where the line is 3" is no greater a change than one that leads to the line being 2" in length. This is because both of these changes amount to merely changing the length of the line to accommodate the counterfactual supposition that it is over 1" long. We can then make whatever other changes are required in order to achieve consistency with the new length of the line.

The second problem with Lewis's theory is its reliance on the notion of similarity itself. Lewis claims that we can make sense of his notion of similarity between possible worlds in some intuitive way, though he thinks that it is essentially vague[6] and context-dependent. He says it is similar to the sort of vagueness that would arise if I were to ask whether Seattle resembles San Francisco more closely than it resembles Los Angeles (Lewis, 1973, p. 92). One's response to this would vary depending on whether one was attaching more importance to the surrounding landscape, the architecture, the dominant industries, the political temper or whatever else. Similarly, in the case of possible worlds, one's response might vary depending on whether one was attaching more importance to the laws of nature that held there or to the particular facts that were true there. However, despite this, Lewis claims that we *do* have a comparative notion of overall similarity, whether it applies to people, cities or possible worlds. Lewis tries to motivate basing his theory on such a vague notion. He says that

> [I] am not one of those philosophers who seek to rest fixed distinctions upon a foundation quite incapable of supporting them. I rather seek to rest an unfixed distinction upon a swaying foundation, claiming that the two sway together independently. The truth conditions for counterfactuals are fixed only within rough limits; like the relative importance of respects of comparison that underlie the comparative similarity of worlds, they are a highly volatile matter, varying with every shift of conflict and interest. (Lewis, 1973, p. 92)

However, this is confused. On the one hand, Lewis realizes that counterfactuals are context sensitive, but, at the same time, he wants to give an account of their truth conditions that works *independently* of contextual factors. Therefore, he claims that counterfactual conditionals have imprecisely fixed truth conditions. Yet, if there is context sensitivity, counterfactuals simply are not true or false independently of relativization to context. Once we recognize this, however, we can see that they can then have precise, fixed, *context-relative* truth conditions.

---

[6] It sounds somewhat odd to say that a notion is "vague." We ordinarily say that words are vague, not notions.

Another important worry is that Lewis begs certain questions about the nature of the vagueness involved here. It could be that the vagueness exhibited by counterfactual conditionals is epistemic in nature. One might argue, for instance, that the reason why we are unsure whether some counterfactual conditionals are true or false is that we are just ignorant about the conditions and causal laws that would support such a conditional. One is not forced to the conclusion that they must be based on a vague relation.[7]

I see no compelling reason to think that we have any intuitive notion of overall similarity between possible worlds *when diverse and competing factors are involved.* One might be inclined to say that there can be no intuitive notion of similarity between possible worlds simply because the notion of a possible world is not one that is entertained by ordinary speakers of English. However, I think that a little more can be said. Clearly, we sometimes employ an intuitive notion of overall similarity, but it is difficult to motivate the claim that there is any such intuitive notion when it comes to possible worlds. We can make sense of a claim such as 'Rossini's operas are more similar to Mozart's than they are to Puccini's'. There may be many factors involved when we compare different music, but this pales in comparison when we consider the number of different factors that would be involved in a comparison of possible worlds. Not only are the factors that we must take into consideration more numerous, but they are also competing and, most importantly, incommensurable. Perhaps we can make sense of questions about more specific similarities between possible worlds than complete overall similarity. For instance:

> Q1: Is world *A* more similar to world *B* than world *C* with respect to the laws that hold there?
> Q2: Is world *A* more similar to world *B* than world *C* with respect to particular facts?

What one finds harder to answer is the following:

> Q3: Is world *A* more similar to world *B* than world *C* with respect to particular facts *and* the laws that hold there?

One can conceive of being able to answer Q3 if the answers to Q1 and Q2 happened to be both 'yes' or both 'no'. Then the similarity concerning laws would not compete with the similarity concerning particular facts. However, when they compete, our intuitions fail us, and judgements of overall similarity between possible worlds *do* seem to involve large numbers of competing factors. The problem, I think, is that the competing factors are incommensurable - there is no common scale on which to compare them.

There is little reason to think that we would have an intuitive concept of something for which we have no actual use (and here I mean *use outside of any philosophical theory*). Given the peculiar results mentioned above, it is difficult to see how the notion of similarity could function in our judgements about counterfactual conditionals. We have seen that it just is not the right sort of relation. Moreover, it is

---

[7] To say that a relation is "vague" is similarly odd to saying that a notion is "vague."

difficult to see how it could have any other good use outside of our judgements about counterfactual conditionals. If this is the case, then a notion of overall similarity between possible worlds could have no use except for the goals and purposes of a theory like Lewis's in which case it is surely just a notion being introduced by Lewis for his theoretical purposes alone. However, it could then not play the role intended for it in providing an intuitive gloss on Lewis's semantics.

Minimal Change Theories

Noticing the problems with Lewis and Stalnaker's similarity relation, other philosophers have attempted to give possible worlds analyses of counterfactual conditionals without using such a relation. Two such theories that I will consider here are those of Pollock (1976) and Nute (1980). Pollock's analysis employs a "minimal change" relation between possible worlds (to be explained below) in place of Lewis's similarity relation. Nute tries to avoid referring to any relations between possible worlds other than those standardly imposed on them, e.g., by modal logicians.

Pollock's analysis of subjunctive conditionals incorporates various elements of both linguistic approaches and possible worlds approaches. Pollock provides an analysis based on possible worlds and an ordering relation upon those worlds. The main difference between Pollock's ordering relation and Lewis and Stalnaker's *closeness* relation is that Pollock's relation only imposes a partial ordering on the worlds whereas Lewis and Stalnakers' ordering is complete. Pollock employs an ordering relation that he calls *minimal change* in preference to Lewis's notion of *overall similarity*. Pollock treats possible worlds as maximal states of affairs and states of affairs as sets of possible worlds. His account quickly becomes extremely complex since he employs many technical notions in order to define precisely the notion of a minimal change.

A central problem with linguistic analyses of subjunctive conditionals is that, in order to be plausible, they are ultimately forced to give truth conditions for subjunctives not just employing logical laws but also, for instance, causal and psychological laws. If we allow just logical laws in L,[8] many subjunctive conditionals that we think are true will be false according to the linguistic analysis. However, as Pollock points out, laws themselves are fundamentally counterfactual in nature. A law, he says, does not simply tell us that all *actual* pulsars are neutron stars, but also that *any pulsar there could be* would be a neutron star. If we express physical laws in terms of subjunctive conditionals, the linguistic analysis is in obvious danger of circularity. Pollock believes he is able to avoid such circularity by defining laws in terms of inductive confirmation. Central to Pollock's account is the idea of a *basic strong subjunctive generalization*. A basic strong subjunctive generalization is a statement of the form *Any A would be a B*. Pollock draws a distinction between *weak* and *strong* subjunctive generalizations. He claims that all physical laws are *strong* subjunctive generalizations. A weak subjunctive generalization, unlike a strong one, depends on a contingent matter of fact for its truth. Consider, for instance, the generalization 'Any human who drank from the container marked "bleach" would be sick'. It is a contingent matter of fact that the container in question actually contains something that is harmful to humans rather than, for instance, lemonade. It is also a contingent matter of fact that humans have a physical constitution that makes bleach harmful to them. We could think of a way in which someone could drink from the container and not become sick, e.g., by washing out the bleach first, but we believe that the generalization is true because this is not actually going to happen. Pollock says that such weak generalizations are not about all *physically possible* persons who might drink from the container, but all *actually possible* persons. A strong subjunctive generalization expresses a lawlike connection that does not depend on such contingent matters of fact, e.g. 'Any creature who had a physical make-up like ours who drank bleach would be very sick'. There is a further distinction between *basic* and *non-basic* subjunctive generalizations. Basic subjunctive generalizations are those ones that are confirmed inductively by their instances. Non-basic generalizations are then derived logically from the basic ones. Pollock (1976) gives an account of inductive confirmation and an account of the way that non-basic generalizations are derived, though I will not discuss the details here.

---

[8] L is the law sentence, or conjunction of law sentences that is supposed to ensure that the antecedent of a counterfactual conditional entails its consequent.

Another important notion in Pollock's analysis of counterfactuals is that of a *simple state of affairs*. Pollock defines a simple state of affairs as one that is logically simple. He thinks that some concepts are complex in the sense that they literally *are* conjunctions, disjunctions etc., not simply *equivalent* to conjunctions and disjunctions. A simple state of affairs, he says, exemplifies a concept or the negation of a simple concept at a specified time. This sort of talk is somewhat difficult to decipher. Nute (1984) interprets Pollock as saying that a state of affairs is simple if it can be known non-inductively to be the case without first coming to know some other state(s) of affairs which entail(s) it (citation).

Pollock says that we are interested in the set of worlds that might be actual (from the point of view of the actual world) if *P* were to obtain. Pollock claims that a definition of those worlds that does not proceed in terms of counterfactuals will be enough to provide an analysis of counterfactuals.

P is counter-legal at a world α just in case P cannot obtain at any world having the same laws as α. Pollock, *contra* Lewis, accepts the principle of *legal conservatism*, which says that

> if P is not counter-legal at α then the laws in the nearest P-worlds to α must have the same laws as α, and if P is counter-legal at α then the laws in the nearest P-worlds must be changed as little as possible from the laws in α.

Legal conservatism puts a constraint on the laws that hold in the nearest P-worlds, but, in order to provide a complete analysis of simple subjunctives, Pollock also needs to put constraints on the simple states of affairs that hold in these worlds. He does this by appealing to a phenomenon that he calls *undercutting*. In order to define 'undercutting,' we need to introduce some additional terminology. A state of affairs P has *historical antecedents* if there is a set of true simple states of affairs Γ such that all times of members of Γ are earlier than the time of P, and Γ nomically implies P. A *nomic pyramid* is a set of states of affairs which contains the historical antecedents of each of its members. If P has historical antecedents in α then P is said to be *grounded* in α. A nomic pyramid is said to trace out the causal history of each of its members until it comes to an ungrounded state of affairs, i.e., a state of affairs with no causal history. In cases of causal overdetermination, a state of affairs will have a plurality of merging causal histories. P is said to *undercut* Q just in case, for every set Γ of true states of affairs such that Γ is a nomic pyramid and $Q \in \Gamma$, Γ nomically implies ~P. The idea is that in revising the set of true simple states of affairs *S* in order to accommodate the antecedent of a subjunctive *P*, we should minimize the deletion of members of *S* which are not undercut by P. Pollock defines the undercutting constraint formally, but it is quite technical, so I will not describe how he does this here.

To summarize Pollock's analysis, when picking the nearest P-worlds to a world w, we must obey two constraints. First, of course, eliminate those worlds where P is false. Second, eliminate those worlds which make more than minimal changes to the strong basic subjunctive generalizations of w. Third, eliminate those worlds that make more than minimal changes to the non-undercut states of affairs of w.

Lewis's counterexample for the Limit Assumption (the one involving the line) is not a counterexample on Pollock's account. For Pollock, there is no undercutting involved in the example. Given that the line is actually 1" long, for any *small* value of x, there will be no more undercutting if the line is 1+x" long than for any other small value, so Pollock says that there *will* be worlds minimally changed from the actual world to make the line 1+x" long.

Minimal change, however, is no more likely to be the relation involved in our judgement of simple subjunctives than the similarity relation is. Pollock's use of counterexamples against the similarity analysis causes one to wonder whether other kinds of counterexamples might be used against Pollock's own minimal change analysis. I think they can. Suppose that there are five graduate students in a philosophy department (Ann, Bob, Carol, Dan and Emma) and they are all twenty-five years old. According to Pollock's analysis, it would then be true that

> if (*P*) the average age of the graduate students were twenty-six, then (*Q*) one graduate student would be thirty years old and the rest would be twenty-five.

Clearly, we have absolutely no reason for accepting such a statement. According to Pollock's analysis, when constructing the nearest *P*-worlds to the actual world, the goal is to first minimize changes to the basic laws (i.e., the basic, strong subjunctive generalizations) and then to minimize changes to the non-undercut simple states of affairs. The above conditional is not a counterlegal, so the requirement about minimizing changes to basic laws does not apply. We are therefore simply required to make minimal changes to the non-undercut states of affairs in order to accommodate the antecedent *P*. If I understand correctly, *P* does not undercut any of the following:

> (*A*) Ann is twenty-five years old
> (*B*) Bob is twenty-five years old
> (*C*) Carol is twenty-five years old
> (*D*) Dan is twenty-five years old
> (*E*) Emma is twenty-five years old

P is not nomically incompatible with every nomic pyramid containing *A*, i.e., the average age of the students being twenty-six does not nomically imply the falsity of any of the historical antecedents of Ann's being twenty-five years old. The students ages having a certain average has absolutely no nomic connection with how Ann happened to reach the age of twenty-five (we can assume that the students take correspondence courses and have never come in contact with each other, in which case there is no causal interaction between them). Therefore, P does not undercut A. For similar reasons, P does not undercut *any* of B, C, D or E. However at least one of A, B, C, D or E must be false in order to accommodate P. Since we are required to make minimal changes to non-undercut states of affairs, we should change as few of the above as we can, and it turns out that we can change just one, by making one of the students thirty years old in our P-world. Doing it any other way would mean changing more than one of the above, so it would not result in a minimal change. Therefore, the nearest P-world must be one in which one of the students is thirty years old and all the rest are the same age.

On reflection, an analysis that involves the idea of making minimal changes to a set of (simple) states of affairs does not seem promising right from the start. At least, there is no reason to think that it leads to an analysis that is any more plausible than one based on similarity. Surely, when we are assessing a hypothetical counterfactual situation, the *number* of states of affairs that we have to change in order to accommodate our counterfactual supposition is not what we are interested in. The possible worlds theorist either needs to find some other way of picking out the "closest" antecedent worlds or abandon the notion of "closeness" altogether. The latter strategy, which appears the more promising, is the one employed by Donald Nute (1980).

Nute's Analysis

Nute claims to provide an analysis of a process that he calls 'hypothetical deliberation'. This, he claims, is the process we engage in when we evaluate a conditional and decide whether it is true or false. Nute bases his analysis on the evaluation of conditionals through a search for counterexamples. He claims that we have grounds for saying that a conditional 'A > B' is false if we can imagine a *reasonable* situation in which A would be true and B would be false. Most of the details of the account are concerned with specifying precisely what he means by a *reasonable* situation. We can imagine all sorts of strange situations, but very few of them will be considered reasonable counterexamples to our counterfactual conditional claims. For instance, suppose that there is a candle on the table in front of me right now. Suppose I wish to assess the conditional claim 'If I were to light the candle, it would burn with an orange flame'. I can certainly imagine a situation in which I light the candle and it starts to burn with a green flame. This certainly shows that my lighting the candle does not *logically entail* that it will burn with an orange flame, but it is not a counterexample to the conditional mentioned above.

Nute describes the process of hypothetical deliberation before going on to provide a formal semantics based on class selection functions. He claims that the process of hypothetical deliberation proceeds as follows:

> *Step 1:* Consider all situations in which the antecedent 'A' of the conditional under question is true. In particular, what actually happens is one of the situations which must be considered if A is actually true. Call this class of situations S(A)
> *Step 2:* Eliminate from S(A) all situations which do not satisfy the relevant conditions of reasonableness. Call the resulting class of situations R(A)
> *Step 3:* Eliminate from R(A) all situation s in which the consequent B of the conditional under question is true. The resulting class of situations is R(A) - S(B). Class R(A) - S(B) is a class of counterexamples to the conditional being considered.
> *Step 4:* Reject A > B *on this occasion* if R(A) - S(B) is not empty. (Remember that

> the standards of reasonableness may change from occasion to occasion).
> *Step 5:* Accept A > B on this occasion if R(A) - S(B) is empty. (Nute, 1980, p. 13)

Nute's most important task is to elucidate the notion of *reasonableness*. Although he thinks that what hypothetical situations count as reasonable counterexamples is strongly context dependent, he also believes that there are certain common features that any legitimate notion of reasonableness must share. These are as follows:

> *Reasonableness condition 1:* If it is impossible for sentence B to be true, then R(A) does not contain any situations in which B is true, i.e., R(A) union S(B) is empty.
> *Reasonableness condition 2:* R(A) is empty just in case A is impossible.
> *Reasonableness condition 3:* R(A) is empty just in case R(~~A) is empty)
> *Uniformity Principle:* If R(A) is empty, then there is no situation R(B) in which A is true, i.e., R(B) union S(A) is empty. This principle can be derived from the first two reasonableness conditions.

Like Lewis, Nute employs a class selection function to capture his intuitions about what counts as a reasonable counterexample to a conditional claim. Some intuitively obvious formal constraints are put upon this function (motivated by the above reasonableness conditions), but these constraints do not rule out very much.

There are no obvious internal problems with Nute's theory, largely because it does not employ any problematic similarity or minimal change relation like most other possible worlds based theories. As far as Nute's description of the process of hypothetical deliberation is concerned, I am mostly in agreement. There is little doubt that something very similar to the process Nute describes occurs when we think about these conditionals. I also agree that, on most occasions, we need to understand the contexts in which they are uttered in order to interpret them properly.

If there is no obvious internal problem with Nute's theory, as there is with Lewis's, Stalnaker's and Pollock's, why am I reluctant to accept it? There are two reasons. First, the theory relies upon the traditional method of individuating antecedents and consequents which, as I argue in Chapter 1 is problematic and completely ignores the tensed verbs that appear in these conditional constructions. Second, the theory relies on an ontology of possible worlds. All other things being equal, a theory that does not rely on such an ontology (even one that treats possible worlds as sets of sentences) and which explains the phenomena under question equally well is preferable. I believe such a theory is possible, as I will show in the next few chapters.

I see little reason to believe that the process Nute describes has anything to do with the *meanings* of these expressions. As I will argue, this sort of deliberative process occurs when we evaluate of whether it is appropriate to assert these conditionals, but the hypothetical situations we imagine have little to do with their meaning. Rather, the considerations that Nute identifies reflect pragmatic considerations rather than semantic ones. That is this so is suggested by the difficulty we have seen that attaches to the attempt to give any precise rule determining the import of an utterance of a conditional in the relevant class. For if the import depends upon a conversational implicature that must take into account in the context of utterances assumptions about the speaker's beliefs and shared background knowledge, we should not expect any rule to be forthcoming. The difficulty we have encountered is a symptom of the fact that the import is pragmatically determined. The facts that are relevant to interpretation on an occasion are open ended, since it is a matter playing off of what participants in a talk exchange take to be common knowledge, as well as the underlying semantic structure of the discourse.

Summary

Even if we ignore their unwarranted assumption that there are two distinct semantic categories of conditionals, most possible worlds theories run into trouble. This is perhaps not surprising, since they take what is a pragmatic distinction and try to explain it as a semantic one. We should not expect to be able to give a precise formal analysis of a distinction between conditionals that cannot be specified independently of the context in which they are uttered. The contextual factors that contribute to the appropriateness of an utterance of a conditional are likely to be so diverse that they cannot be precisely described using notions like "overall similarity" and "minimal change." This is not to say that notions such as these cannot help us when thinking about conditionals and deciding when it is appropriate to assert them. In fact, it seems that thinking about possible worlds will often help us to decide whether it is appropriate to assert a so-called counterfactual conditional. However, because there is no legitimate semantic distinction between conditionals such as these theories presuppose, these account worlds cannot tell us anything about the meaning of these conditionals.

CHAPTER 5
THE SEMANTICS OF TENSE

Since my analysis of so-called subjunctive conditionals depends heavily on their status as tense expressions, it is important that I provide a semantics for those expressions. The approach I use here treats tense expressions as indexical devices for referring to times at which events take place or states of affairs obtain (I will henceforth refer to this approach as "indexical semantics"). This approach originally goes back to Frege and Russell but has recently been developed in detail by Ernest Lepore and Kirk Ludwig. In this chapter, I outline my preferred analysis of tense expressions and contrast it with the more traditional approach that usually goes under the name of "tense logic."

Truth-theoretic semantics

Following Lepore and Ludwig, I develop my semantics for tense expressions within the context of a Davidson-style truth-theoretic semantics for a natural language. Davidson was the first to suggest that a Tarski-style theory of truth might be adapted to produce a theory of meaning for a natural language. I will not try to defend truth-theoretic semantics against all objections, and I do not even claim that is the only variety of semantics that will work with my account of conditionals. However, Lepore and Ludwig are right that "truth-theoretic semantics provides a particularly illuminating standpoint from which to discuss issues about the semantics of tense, and their relations to thoughts at, and about, times" (Lepore and Ludwig, 2000, p. 1).

An important virtue of truth-theoretic semantics is that it enables us to give a compositional meaning theory for the language under analysis. According to Lepore and Ludwig, a compositional meaning theory for a natural language L should

> (R) entail, from a specification of the meanings of the (finitely many) primitive expressions of L, all true sentences of form (M),

where (M) is as follows:

> (M) For any speaker s, time t, sentence Φ of L, Φ $\text{means}_{[s,t,L]}$ that p. (Lepore and Ludwig, 2000, p. 5)

'Φ $\text{means}_{[s,t,L]}$ that' is to be read as 'Φ understood as it would be in L if spoken by s at t means that.' Relativizing meaning to speakers and times is necessary if we are to give a proper account of demonstratives and, as we will see, a proper account of tense expressions. We can use a Tarski-style truth theory as a means of producing the required theorems of form (M). We will say that a truth theory for a natural language meets *convention T**[1] only if it entails every instance of the following schema:

> (T*) For any speaker s, time t, sentence Φ of L, Φ is $\text{true}_{[s,t,L]}$ iff p

(where a metalanguage expression which interprets Φ replaces 'p'). As one would expect, 'Φ is $\text{true}_{[s,t,L]}$ iff' is to be read as 'Φ understood as it would be in L if spoken by s at t is true iff.' I will call instances of (TLL) *TLL-sentences*. It is intuitively clear that the relation between the sentence that replaces 'Φ' and the metalanguage sentence that replaces ‘p’ in a TLL-sentence is the same relation that holds between the sentences that replace ‘Φ’ and ‘p’ in a sentence of form (M). Thus, we have the following schema:

---

[1] I call it 'Convention T*' because Tarski's Convention T, on which it is based, was never intended to apply to natural languages.

> (S) For any speaker s, time t, sentence Φ of L, if (Φ is $\text{true}_{[s,t,L]}$ iff p) then (Φ $\text{means}_{[s,t,L]}$ that p).

(S) enables us to use a TLL-sentence for s in order to specify its meaning. I will use this schema to provide a truth conditional semantics for tense expressions, and in Chapter 7, I will use it in my analysis of so-called subjunctive conditionals.

The Quantificational Treatment of Tense

Once again following Lepore and Ludwig, I treat tense as an indexical device. Past and future tenses in particular involve what Lepore And Ludwig call *indexically restricted quantifiers*. The indexical element picks out an utterance time relative to which temporally bound states and events are quantificationally located. It seems natural to treat tensed verbs as quantifying over event times. Most philosophers have little trouble accepting accounts of modal expressions that analyze them as quantifying over possible worlds, so we should have even less trouble accepting an account of tense expressions that quantifies over moments in time. Of course, there is an underlying assumption that time is real. If time is not real and the indexical account of tense is correct, then almost everything we say is false.

It is essential that our semantics for tense be expressed in a context insensitive language. This ensures that the semantic theory is interpersonal, in the sense that anyone can potentially state it. As Lepore and Ludwig point out, any theory with indexical or other context sensitive referring terms cannot meet this condition, since two people in relevantly different contexts would fail to say the same thing. This causes some difficulties for us here, because tense expressions are themselves context sensitive. When one uses the future tense, for instance, one uses it to refer to a time in the future *relative to the time of one's utterance.* One might say that an utterance of 'Mary will leave Bill' means that, *at some time after the time of the speaker's utterance,* Mary will leave Bill. However, this analysis is circular, because it reuses the very same tense that it tries to explain.

In order to give an adequate semantics for tense expressions, we need to relativize utterances containing tense expressions to the times of their utterances. This entails treating tense expressions as indexical devices. Where, 'y' is any indexical, I use '$\text{ref}_{[s,t]}(y) = x$' to abbreviate 'the referent of y as potentially used by s at t = x'. In order to satisfy the requirement of context insensitivity, our metalanguage includes tenseless verbs expressing relations between objects and events and times. I use the present tense form with a subscripted variable indicating temporal argument place, e.g., 'loves_t' in the metalanguage corresponds to 'loves' in English, relativizing it to time t. I use 't*' to refer to 'speaker time', i.e., the time of a speaker's utterance. More formally, we say that, for any speaker s, time t, $\text{ref}_{[s,t]}$('t*') = t.

Natural languages lack explicit markers for temporal variables, without which it is difficult to construct a semantics for tensed expressions. Thus, my truth theory is given for English*, which extends English by 't*' and tenseless verbs with explicit markers for temporal argument places corresponding to English verbs. '$x < y$' and '$x > y$' will be used to mean 'x is earlier than y' and 'x is later than y' respectively.

It is not difficult to modify this analysis of tense expressions to provide an account of temporal adverbials. For example, we can specify a reference clause for 'tonight' as follows:

For any speaker s, times t, t_1, such that t_1 = the night of t, $ref_{[s,t]}$('tonight') = t_1

The truth conditions for a sentence such as 'Joe will work tonight' can then be represented as follows:

'Joe will work tonight' is $true_{[s,t]}$ iff [There is a t_1: $t_1 > t$ and $t_1 = ref_{[s,t]}$('tonight')] (Joe $works_{t1}$)

Table 1 contains treatments for verbs in some simple and perfect tenses. The motivation for these treatments should be obvious to the reader.

Table 1: Lepore-Ludwig treatments of past tenses.

Tense	Example	Treatment
Simple Past	Bill left	[There is a t_1: $t_1 < t^*$] (Bill $leaves_{t1}$)
Simple Future	Bill will leave	[There is a t_1: $t_1 > t^*$] (Bill $leaves_{t1}$)
Past Perfect	Bill had left	[There is a t_1: $t_1 < t^*$] ([There is a t_2: $t_2 < t_1$] (Bill $leaves_{t2}$))
Present Perfect	Bill has left	[There is a t_1: $t_1 = t^*$] ([There is a t_2: $t_2 < t_1$] (Bill $leaves_{t2}$))
Future Perfect	Bill will have left	[There is a t_1: $t_1 > t^*$] ([There is a t_2: $t_2 < t_1$] (Bill $leaves_{t2}$))

Of special concern will be the treatment of the past-shifted future modal 'would' in both simple and perfect future tenses. Consider the contrast between (1) and (2).

(1) A child was born who would be king.
(2) A child was born who will be king.

In the second of these, the reference time for the tense in the relative clause is the present, that is, this says a child was born who at some time in the future of the time of utterance will become king. (2) is false if no child was born who will in the future become king. So if a child was born who became king in the past, but none born in the past will become king, (2) is false. In contrast, (1) would be true in those circumstances, for the 'would' in (1) is interpreted as future relative to the event time expressed by the main verb. Thus, we would represent semantically (1) as in (1a), in contrast to the semantic representation of (2) in (2a).

> (1a) [There is a t1 < t*][there is a t2: t2 > t1](a child is$_{t1}$ born and he is$_{t2}$ king).
> (2a) [There is a t1 < t*][there is a t2: t2 > t*](a child is$_{t1}$ born and he is$_{t2}$ king).

The crucial difference here is that the reference time for the verb in the subordinate clause is determined by the event time of the main verb in (1a), but indexes to speaker time in (2a). Other examples of the future in the past are given in (3) and (4).

> (3) Jack said that he would do it, but he hasn't yet.
> (4) No one who saw the man's confession would ever regard him as quite trustworthy again.

Clearly in (3) the reference time of the verb in the complement clause of the indirect discourse report is determined by the event time of the main verb. In (4), the reference time for the future modal is determined by the event time of the verb in the nominal of the noun phrase, and similarly for the past-shifted future perfect, illustrated in (5).

> (5) It was June 5th, and by the end of the next day a beachhead would have been established on the Normandy coast, though at great cost.

It is hard to find uses of the past shifted future or future perfect in which the argument place fixed indexically in the unshifted future modal is not intuitively bound or determined by the tense of another verb in the sentence or in the surrounding discourse. We can represent the patterns in which we find the past shifted future and future perfect as follows:

Table 2: Treatment of relativized future modals.

Tense	Example	Treatment
Future-in-the-past	Bill would leave	[There is a t1: t1 <t*] [there is a t2: t2 > t1] (Bill leaves$_{t2}$)
Future-perfect-in-the-past	Bill would have left	[There is a t1: t1 <t*] [there is a t2: t2 > t1] . . . . . . [there is a t3: t3 < t2] (Bill leaves$_{t3}$)

The above tenses also occur in counterfactual conditionals. For instance, Ernest Adams' example, 'If Oswald had not killed Kennedy then somebody else would have', seems to have the past perfect in the protasis and the future-perfect-in-the-past in the apodosis. The sentence 'Oswald had not killed Kennedy' clearly contains a past perfect verb. It would indeed be odd if, when this construction is embedded in a conditional sentence, its meaning changes so completely as to, for instance, say something about possible worlds. Thus, I claim that there is nothing special about conditionals such as this other than that they are tensed and that we should analyze them using similar logical apparatus to that used for the above sentences.

Tense Logic

The indexical approach to tense has not received the attention it deserves. Most of the work that philosophers and logicians have done on the analysis of tense expressions has come under the auspices of 'tense logic'. Tense logic was not originally conceived with the analysis of the tenses of natural language in mind but as an attempt to introduce tense expressions into the propositional calculus. The primary difference between tense logic and the indexical approach that I endorse here is that tense logic analyzes tense expressions as sentential operators rather than as quantifying over times.

Many regard A.N. Prior as the father of modern tense logic, though Prior himself cites J.N. Findlay. Findlay believed that the calculus of tenses should have been included in the development of modern logics since our conventions with regard to tenses are so well worked out that we have practically the materials in them for a formal calculus. It is certainly true that most work in tense logic aims to provide a formal calculus of tenses rather than a true semantics for tense in natural languages.

Standard tense logics introduce non-truth functional sentential operators such as the following (though there are many other variants):

> 'PA' to be read as 'it was/has been the case that A' or '"A" was true'
> 'FA' to be read as 'it will be the case that A' or '"A" will be true'
> 'JA' to be read as 'it is now the case that A' or '"A" is true now'

The 'it . . . the case that' reading seems preferable since, the other reading treats the tense operators as expressions for talking about sentences rather than events or states. These operators give us the following equivalences, reminiscent of equivalences between the operators of modal logic:

> FA =df. ~G~A, i.e., 'It will be the case that A' is equivalent to 'It will not always be the case that not A'
>
> PA =df. ~H~A, i.e., 'It has been the case that A' is the equivalent to 'It has not always been the case that not A'

I am not particularly interested in the details of tense logics here, so I will not formally define any systems of tense logic. I will mainly be concerned with the idea of treating tense expressions as sentential operators as opposed to indexical devices referring to times.

It would be unfair to treat tense logic and the indexical approach to tense as rival approaches in natural language semantics. Tense logic does not really attempt to give a semantics for English. It is primarily concerned with introducing tense expressions into formal languages, its main goal being to show how tense expressions interact with one another. Dov Gabbay (1976. P. 167) claims that traditional tense logic is mainly concerned with axiomatizing sets of valid statements for various different models of time. For instance, tense logic might tell us interesting differences about how tense expressions interact when we switch from the assumption of branching time to the assumption of non-branching time. It must still be the responsibility of the tense logician to explain to us what these differences in models of time really amount to. Though it might be mathematically interesting, it is not particularly philosophically useful to model different kinds of "time" by imposing different relations on a model structure while providing no natural interpretation of what those model structures are suppose to represent.

Comparison of the Two Approaches

Having summarized the two approaches to the analysis of tense, we can ask what is the relation between work in tense logic and the indexical semantics for tense that I prefer to use in this dissertation. Why, in particular, have I chosen to treat tense expressions as indexical devices rather than sentential operators?

The main reason why I have not chosen to give a semantics for tensed sentences in English in the language of tense logic is that tense logic is simply incapable of performing such a task. As I have noted, a semantics for English must itself be given in a metalanguage which is not context sensitive, or at least which does not use context sensitive expressions in giving the satisfaction conditions for English expressions in an interpretive truth theory. However, standard tense operators, such as 'F', 'P', 'J', as well as others which have been suggested, are interpreted so that they are evaluated only relative to times, and sentences of the form 'FA', 'PA' and 'JA' may receive different truth values evaluated at different times. If one were to try to use them to interpret natural language sentences, one would treat utterance time as the relevant time of evaluation. Hence, these operators are not context insensitive, and, therefore, not suitable for use in an interpretive truth theory for English. They are expressions in a language that is like English in respect of having in effect tense devices. This comes out in the standard interpretations of the tense operators since 'F', 'P', and 'J' are interpreted along the lines of 'It will be the case that', 'It was the case that', and 'It is the case that'.

Of course, tense logicians have provided semantics for their logics. Truth values of well-formed formulae of tense logic are defined using "tense structures" (it is not necessary for me to go into the details here). Since these semantics define truth for tense logics, we might try to use them to provide a truth–theoretic semantics for natural language expressions. However, such an exercise is, at best, pointless. It would involve a two-step process of first translating sentences of ordinary English into sentences in the language of tense logic, and then giving them a semantic analysis. A semantics for natural language expressions that does not require us to first regiment those expressions in an impoverished formal language is surely preferable. Moreover, it is not altogether clear whether we are able to adequately express the meaning of all sentences in English using the limited language of tense logic. I return to this question later.

Indexical semantics for tense expressions allows us to analyze every expression that we can express using a tense logic. One way of proving this is to show how the language of tense logic can be interpreted by indexical semantics. The interpretations usually intended for the tense operators in tense logic are given in English. The right interpretation of the tense operators then is given by the proper semantics for their English interpretations (though there is some variation among tense logicians about what the correct English interpretations are said to be).

The languages of tense logic usually take atomic sentences not to themselves contain any tensed verbs. However, atomic sentences are evaluated with respect to times. Usually, no attention is given to how to understand this, but rather a function from sentences and times to truth-values is assumed. Let us however suppose that the proper way of further articulating this is to suppose that the atomic sentences contain predicates with an argument place for time. In this case, each of the standard tense operators can be given interpretations which make them devices which do the same work as verb inflection or modification for the past, present or future tenses.

Let us suppose we are working with a language with atomic sentences and formulas and tense operators and the usual logical connectives, but let us suppose for simplicity that the language contains no quantifiers (excepting as it turns out the tense operators). The extension to a language with quantifiers will be straightforward. Let us represent an atomic sentence without temporal operators but with a temporal argument place for time as ‘A_t’. Then we can give the following satisfaction clauses for ‘F’, ‘P’ and ‘J’ for atomic sentences where we quantify over utterance times (we suppose for simplicity that no expressions of the language require relativization to speaker, though this obviously could be accommodated).

(6) For all t, f sat_t ‘FA_t’ iff [there is a t1: t1 > t] A_{t1}
(7) For all t, f sat_t ‘PA_t’ iff [there is a t1: t1 < t] A_{t1}
(8) For all t, f sat_t ‘JA_t’ iff [there is a t1: t1 = t] A_{t1}

We assume the metalanguage embeds the object language here.

This is in effect to treat ‘F’, ‘P’ and ‘J’ as abbreviations of restricted temporal quantifiers with an argument place that, if not otherwise controlled, indexes to utterance time. To handle iterations of tense operators, we will let ‘T’ range over complex formulae with one or more tense operators. (9)-(11) give their satisfaction conditions recursively.

(9) For all t, for all T, f sat_t ‘FT_t’ iff [there is a t_1: $t_1 > t$](f sat_{t1} T)
(10) For all t, for all T, f sat_t ‘PT_t’ iff [there is a t_1: $t_1 < t$](f sat_{t1} T)
(11) For all t, for all T, f sat_t ‘JT_t’ iff [there is a t_1: $t_1 = t$](f sat_{t1} T)

Clauses (6)-(11) give a complete recursive semantics for the simple tense operators for this simple language. Let us add (12):

(12) For all t, for all A, f sat_t ${}^{\Pi}A(X,t)^{\cdot}$ iff f sat_t ${}^{\Pi}A(f(X),t)^{\cdot}$
[Is this right? I fixed it a bit from the email you sent me. I always get muddled by satisfaction clauses!]

where ‘X’ stands for the free variables in ‘A’ besides ‘t’ and ‘f(X)’ for the assignment by f to those variables. Also, in (9)-(11), we will let T be any formula of our logic. (6)-(12)

supplemented with standard satisfaction clauses, give the form of a completely general semantics for a language with tense operators.

I now turn to the question of whether all natural language sentences can be paraphrased correctly in a tense logic. The first thing to say is that obviously any given system of tense logic may not have adequate expressive powers to paraphrase tense devices in English. The question is whether something in the general family of tense logics could be adequate. If we put no restrictions on the number of tense operators we allow to be introduced with special interpretations, the answer to this question is that there appears to be good reason to think that an adequate tense operator notation could be devised. The reason is that on our interpretation above, tense operators are treated as restricted indexical quantifiers. Our semantics for tense, if adequate, represents tense as contributing, expect in the case of the present tense, quantifiers. The aim would be to work backwards to an adequate tense operator notation.

However, we can expect to encounter some difficulties. First, we need an adequate way of representing the perfect tenses. The simple tense operator language we have just examined is not adequate. It can handle certain more complex tense constructions. For example, 'FP' and 'PP' can be used to represent the future perfect and past perfect. It is also not too difficult to treat the future perfect in the past. The contrast between (13) and (14)

(13) A child was born who will be king
(14) A child was born who would be king

can be treated as a matter of different scope relations, and represented as

(15) P(A child is born) and F(he is king)
(16) P(A child is born and F(he is king))

The analyses in (15) and (16) do not make mention of any temporal argument place. We would have to add one if we were to give an analysis of (15) and (16) using our indexical approach. If we can use 'PP' for the past perfect and 'FP' for the future perfect, we might expect to be able to use 'JP' (or maybe 'PJ') for the present perfect. Unfortunately, the present perfect does not seem to be equivalent to either 'PJ' or 'JP'. Consider

(17) He has been practicing
(18) He was practicing

Sentence (17) is true if he is practicing at the time of utterance, though (18) might be false in that case. However, when (18) is true so is (17). Consider (19), (20) and (21):

(19) PJ(He is practicing)
(20) JP(He is practicing)
(21) P(He is practicing)

Sentence (19) will be true in exactly the same circumstances as (20) and (21). So, we need a new tense operator for the present perfect if we are to be able to distinguish it from the simple past.

A second problem concerns the grammar of iterated tense in tense logics. Suppose we want to translate the sentence 'FP(Joe worked)' into English. If we read 'FA' as '"A" will be true' we get:

'"Joe worked" will be true' was true.

If we read 'FA' as 'It will be the case that A', we get:

It will be the case that it was the case that Joe worked.

Both of these translations are of dubious grammaticality. Clearly some additional reconstruction of sentences is required when translating tense logic into English if that translation is to be adequate.

Another complication is that we need a way of representing appropriately tense interaction and independence in a sentence. Consider (19)

(19) John believed he would be here.

How we treat this will depend on the semantics we give for give for attitude sentences. A natural paraphrase would be (20).

(20) P(John believes F(John is here))

In order to treat this properly, we need to know what to do with the complement sentence. In order to do that, we need a semantics for belief contexts, which is something that I am unable to produce here. Whatever form that semantics takes, it must be able to track the interaction between the verb in the main clause and the verb in the complement clause such that it represents John's belief as being a belief about the appropriate time.

Another difficulty in translating English into the language of tense logic is that we need a way of showing how temporal adverbials can be incorporated into a tense operator language. Consider (22).

(22) John slept at midnight.

What this calls for is a way of representing modifiers of the tense operators themselves. We can simply adopt the expedient of attaching the modifier to the appropriate operator in brackets after it. Thus, (22) is paraphrased as (23).

(23) P[at midnight](John sleeps).

Then, we must give an appropriate semantics for the modified operators. In this case, (24) gives us what we require:

> (24) For all f, t, f sat_t 'P[at midnight]$(\text{John sleeps})_t$' iff [there is a t_1: $t_1 < t$ & $t_1 =$ ref(midnight,t)](f sat_{t1} 'John sleeps_t').

Obviously, there are many more complexities to deal with, but there seems no reason in principle we could not devise an appropriate notation for a tense logic with an appropriate semantics that would have the same expressive powers as English. The simple reason for this is that if we have an adequate semantics for English it already provides us with an adequate semantics for our tense logic provided that we enrich it enough to capture all the features of natural language tense devices. The result is simply a language that in its syntax and semantics is able to express just what we can express in English using tensed devices.

Summary

In this chapter, I have discussed two approaches to tense logic. The first approach treats tense expressions as indexical devices for referring to times at which events take place or states of affairs obtain. This approach has a great deal of intuitive plausibility, and it is very useful for analyzing tense expressions in ordinary English. The second approach, which I have referred to as the "tense logic" approach, treats tense expressions as sentential operators. Tense logic requires us to translate English sentences into sentences containing these operators (except in those rare cases where the English sentence contains the operators already).

Tense logics are not suitable to provide a semantics for English sentences. Rather, we should treat them as alternative tensed languages, into which English sentences may be paraphrased with more or less success depending on the richness of the language. They cannot provide a semantics for English sentences precisely because they can paraphrase the tense devices of English, and these must be eliminated in giving a semantics for them in order to fulfill the requirement that the language of our theory must be context insensitive.

Tense logics are useful because they enable us to define formal languages containing tense expressions. We may take this ability for granted now, but modern logic had existed for over fifty years before Prior made the breakthrough. Using these formal languages, we can construct axiom systems to model the different ways tense expressions interact with each other. However, we are unable to tell whether a particular tense logic describes these interactions correctly unless we already have some notion of how tense is supposed to work in English. Therefore, tense logics actually presuppose an adequate semantics for English expressions rather than providing such an analysis themselves.

These considerations suggest that a semantics based upon the indexical approach to tense will be the most fruitful in providing an account of the use of tense expressions in various conditionals. I therefore employ such a semantics in Chapter 7, where I offer a comprehensive account of so-called subjunctive conditionals.

CHAPTER SIX
INDICATIVE CONDITIONALS

In this chapter, I explain and defend the theory of indicative conditionals that underlies my analysis of so-called counterfactuals. Though this is primarily a work on so-called counterfactual conditionals, it would not be complete without an analysis of indicative conditionals. Many authors claim that so-called subjunctive and indicative conditionals are closely related and interchangeable in many instances. This requires an explanation. More importantly, my own analysis of so-called counterfactuals treats them as quantified indicative conditionals, so I am obliged to give an account of the meaning and assertability conditions of indicatives. The theory I offer has much in common with those of H.P. Grice and Frank Jackson, since it is based on the claim that the indicative conditional '$(A \rightarrow B)$' and the material conditional '$(A \supset B)$' are logically and semantically equivalent. I accept the broadly Gricean defense of this thesis which sees the standard import of these conditionals as a matter of the interaction of their truth-functional semantics with conversational dynamics. I show how to extend this defense to so-called subjunctive or counterfactual conditionals.

Jackson, Grice and The Equivalence Thesis

In his excellent book *Conditionals*, Frank Jackson defines the *equivalence thesis* as the thesis that the material conditional, the general form of which is represented as '$(A \supset B)$', is logically equivalent to the indicative conditional, the general form of which we can represent as '$(A \rightarrow B)$'. As we shall see, I also claim that they are *semantically* equivalent, but Jackson does not. Jackson (pp. 4 – 6) derives the equivalence thesis from three intuitively plausible principles:

> *(TF) The truth-functionality principle*: The standard, familiar, two-valued truth-table accounts of 'not', 'or', and 'and', in their typical uses, are correct. In simple terms, this means that, in ordinary English, negations are true if and only if what they are negations of are false, and they are false if and only if what they are negations of are true. Disjunctions are true if and only if at least one disjunct is true while conjunctions are true if and only if both conjuncts are true.
>
> *(UP) The uncontested principle*: It is logically sufficient for the falsity of an indicative conditional that its antecedent be true and its consequent false.
>
> *(PP) The passage principle*: '(A or B); therefore, (not-A → B)' is valid. 'Not-(A and B); therefore, (A → not-B)' is valid. For example, suppose I am told that either the butler did it or the footman did it. I may infer that if the butler did not do it, the footman did. Suppose I am told that the butler and the footman are not both innocent. I may then infer that if the butler is innocent, the footman is not.[1]

From the truth-functionality principle and the uncontested principle, it follows that 'not-(A ⊃ B); therefore not-(A → B)' is valid. By contraposition, '(A → B); therefore (A ⊃ B)' is also valid. The passage principle, together with the truth-functionality principle, implies that '(A ⊃ B); therefore (A → B)' is valid. Therefore '(A → B)' and '(A ⊃ B)' are logically equivalent, i.e., the equivalence thesis is true.

---

[1] Jackson claims that there is no 'Paradox of Subjunctive Conditionals' of a similar nature to the Paradox of Indicative Conditionals. He argues that this is because the passage principle for subjunctive conditionals is not plausible. Suppose, for instance, that I know that Charles I was executed in 1649 or 1639. I am surely entitled to infer from this that if he was not executed in 1649 then he was executed in 1639, as the passage principle says. However, I am not entitled to infer from this that if he had not been executed in 1649 then he *would have been* executed in 1639. If the passage principle also applied to subjunctive conditionals, this would have to be a valid inference. Though this reasoning seems correct, it relies upon the philosopher's standard procedure for reconstructing the antecedents and consequents of so-called 'subjunctive' conditionals by changing the tense of the verbs that appear in their apodoses and protases. I reject this procedure.

However, controversy is not far away. If the equivalence thesis is true, then the truth-functional validity of 'not-A; therefore $(A \supset B)$' and 'B; therefore $(A \supset B)$' imply the validity of 'not-A; therefore $(A \rightarrow B)$' and 'B; therefore $(A \rightarrow B)$' respectively. This result contradicts a fourth principle which Jackson calls *the principle of the paradoxes of material implication*.

> *(PPMI) The principle of the paradoxes of material implication*: 'not-A; therefore $(A \rightarrow B)$' and 'B; therefore $(A \rightarrow B)$' are invalid inference patterns. In other words, this principle says that material conditionals do not function in the same way as natural language indicative conditionals.

PPMI clearly has some intuitive appeal, and has motivated many people to reject classical truth functional logic as the appropriate logic for indicative conditionals. In particular, it is one of the primary motivations for research into so-called relevant logics. However, a less radical resolution of this apparent contradiction is to reject *PPMI*. Indeed, if we accept that the material conditional is equivalent to the ordinary indicative conditional, this is the only course to take. At the same time, if our position is to be at all plausible, we need to explain why *PPMI* has its initial plausibility. On frequent occasions, we would not assert an indicative conditional even when we believe that the equivalent material conditional is true. We may even be prompted to say that we do not *believe* that many indicative conditionals are true even when we believe that their equivalent material conditionals are true. In particular, we tend not to assert conditionals simply because we know that they have a false antecedent or a true consequent. Consider, for instance, the following conditional:

(1) If Elvis is alive, then he's living in Belgium.

Since we have excellent evidence that Elvis died and has not been resurrected, we know well enough that this conditional has a false antecedent. Therefore, we have good enough reason to believe that the material conditional ‘Elvis is alive $\supset$ Elvis is living in Belgium’ is true (since all material conditionals with false antecedents are true), in which case, according to the equivalence thesis, we have good enough reason to believe that the indicative conditional (1) is also true. However, it would surely be inappropriate, or maybe even misleading, to assert (1), and many people would be inclined to believe that its assertion is false.

The least problematic way to explain these cases is to bring to bear the distinction between semantics and pragmatics. When analyzing the semantics of natural languages, one should always be aware of pragmatic considerations. As I discussed in Chapter 2, in conversation, we use words in all sorts of ways, and frequently we are not able to interpret a person’s utterance by considering its meaning alone. Consider the following example from Grice’s “Logic and Conversation”:

> Suppose that A and B are talking about a mutual friend C, who is now working in a bank. A asks B how C is getting on in his job, and B replies, *Oh quite well, I think; he likes his colleagues and he hasn’t been to prison yet.* (Grice, 1989, p. 24)

Obviously, B is trying to imply something with his remark. He did not simply intend A to take it at face value. Grice suggests that he might wish to intimate that C is the sort of person likely to yield to the temptation provided by his occupation or that C’s colleagues are unpleasant and treacherous people. In any case, it would be perfectly normal for A to ask B what he was implying (though, if A is familiar with C and his working conditions, he might already have guessed). What is clear is that there a distinction between what B implied and what he said. To capture this notion, Grice introduces the noun 'implicature' and the verb 'implicate'.

When you begin a conversation, you enter a sort of contract. You trust your conversational partner to obey the rules of helpful conversation, and he expects you to do the same. This is not to say that the rules are never broken or manipulated, but effective conversational discourse would be impossible without them. Grice attempts to capture these rules with the following (very general) principle:

> *Grice's Cooperative Principle*: Make your conversational contribution as suitable as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. (citation)

Grice identifies four specific categories of conversational maxim under which fall several submaxims. These are:

> *The Category of Quantity*: (1) Make your contribution as informative as is required (for the purposes of the exchange) and (2) [d]o not make your contribution more informative than is required.
> *The Category of Quality*: (1) Do not say what you believe to be false and (2) [d]o not say that for which you lack evidence.
> *The Category of Relation:* Be relevant.
> *The Category of Manner*: (1) Avoid obscurity of expression. (2) Avoid ambiguity. (3) Be brief (avoid unnecessary prolixity). (4) Be orderly. (Grice, 1989, pp. 26 – 27)

The Category of Quantity is the one that will most concern us here. Grice illustrates it with a non-conversational analogue:

> If you are assisting me to mend a car, I expect your contribution to be neither more nor less than is required. If for example, at a particular stage I need four screws, I expect you to hand me four, rather than two or six.

To give a conversational example, if I ask you when you were born, and you say "In the twentieth century," you are violating the Maxim of Quantity by not giving me enough information. Conversely, if you were to tell me the time of your birth to the exact second, you would probably giving me more information than is required.

It is important to note that sometimes people deliberately flout a maxim in order to achieve a certain conversational implicature. For instance, in the bank example, B's response looks like it violates the Maxim of Quantity by not giving enough information about C and his job. However, A understands that B is not deliberately trying to be an unhelpful conversational partner and realizes that he is trying to implicate something that is not explicitly said. It is clear that we can only interpret B's remark properly by considering what it implicates.

The Maxim of Quantity tells us that, in most cases, we shouldn't assert a conditional unless we believe that there is some non-truth functional connection between the antecedent and the consequent, e.g., that the proposition expressed by the antecedent entails or nomically requires the proposition expressed by the consequent (Grice refers to this as the *Indirectness Condition*). If, for instance, one knows that A is false, it would most likely be more informative to assert that A is false than to assert a (true) conditional with A as its antecedent. This is not to say that all assertions of conditionals implicate a connection between antecedent and consequent. The implicature is, at best, conversational since, in many cases, it is cancelable. Grice gives the following example:

> To say 'If Smith is in the library, he is working' would normally carry the implication of the Indirectness Condition; but I might say (opting out) 'I know just where Smith is and what he is doing, but all I will tell you is that if he is in the library he is working.' (Grice, 1989, p. 59)

The cancelability of an implicature is a good indication that the implicature is merely pragmatic and not a product of the meaning or sense of the sentence asserted.

An analysis of conditionals employing Grice's maxims is better able to explain certain cases than other kinds of analysis. Suppose I want to make a cruel joke at the expense of my friend Bob. When asked where Bob is, I reply 'if he's not in the bar, I'm a monkey's uncle'. My interlocutor obviously realizes that I know I am not a monkey's uncle, so he also realizes that my grounds for asserting the conditional could not be that I believe there is any non truth-functional connection between the antecedent and the consequent. He therefore understands that I am "opting out" of obeying the Indirectness Condition and that I am asserting the conditional on the simple, truth-functional grounds that I know that both the antecedent and consequent are false. Thus, he quickly realizes that the purpose of my assertion is to insinuate that Bob is a drunkard who spends most of his time in the bar. A theory claiming that the Indirectness Condition is somehow connected with the *meaning* of an indicative conditional has difficulty explaining this sort of case. If the conditional literally *means* that there is some non truth-functional connection between Bob being in the bar and me having an ape for a relative, it is difficult to see how it could ever be asserted legitimately.

A sentence is appropriately uttered in a conversation if it has a high degree of assertability (unlike truth, assertability is not an all-or-nothing affair). For most simple assertoric sentences, their degree of assertability is equal to their subjective probability of truth. An analysis of the assertability of conditional sentences is somewhat more complex. Here I will argue, similarly to Jackson, that the degree of assertibility of a conditional $(A \rightarrow B)$ is largely determined by the subjective probability of B given A. However, unlike Jackson, I do not claim that the probability of B given A is always sufficient for determining the degree of assertibility of a conditional. On occasion, other factors are involved.

Jackson identifies what I believe to be the most important requirement for the assertibility of indicative conditionals; namely, that under normal circumstances, one should not assert a conditional if it is not *robust with respect to its antecedent*. The idea behind the *robustness* requirement is that you should not assert a conditional if learning the truth of the antecedent would undermine whatever reason you have for thinking that the conditional is true.[2] If my only reason for believing (1) to be true were the falsity of its antecedent, then if I were somehow to learn that Elvis had been resurrected, my confidence in the truth (1) would be undermined. I would no longer have any grounds for believing that the material conditional is true. We must therefore say that, though (1) is true, it is not robust with respect to its antecedent, and consequently, that it has a low degree of assertibility. In addition to the robustness of the entire conditional with respect to its antecedent, the robustness of the consequent to the antecedent is also important. If you believe that the consequent is true, learning that the antecedent is true should not undermine your belief in the truth of the consequent. However, it turns out that the robustness of the entire conditional with respect to its antecedent is sufficient for the robustness of the consequent with respect to the antecedent. The robustness requirement is expressed formally as follows:

---

[2] This requirement has obvious similarities to Mackie's requirement for a universal generalization to sustain a counterfactual claim. See Chapter 3 for a discussion of this.

> (2) It is proper to assert (A → B) when (A ⊃ B) is highly probable and robust with respect to A, that is, when P (A ⊃ B/A) is also high.

If P (A ⊃ B/A) is high, it follows that P (B/A) is also high. Thus, if a conditional is robust with respect to its antecedent, then that conditional's consequent is also robust. The probability involved here is subjective probability. Whether or not a conditional is highly assertible will therefore depend on how the potential utterer assesses the relevant probabilities.

Jackson draws a distinction between the assert*i*bility of conditionals and their assert*a*bility (a distinction also employed by Adams (19xx) and Lewis (19xx)). He says that his robustness criterion relates specifically to the assert*i*bility of conditionals, not their assert*a*bility. "[T]he notion of assert*i*bility we are using is distinct from a wider, more pragmatic notion that I will use assert*a*bility for" (Jackson, 1989, p. 10). There is a distinction, he says, between evaluating a statement for its content, *what* is said, and for the *saying* of it. There is a sense in which I would be warranted in loudly asserting "If I say that the Battle of Hastings was fought in 1066, then I will say something true" even in a silent reading room. At least, I would not have asserted anything misleading or unhelpful. However, there is also a sense that it would not be appropriate because, for instance, it would disturb the people who are trying to read in peace. For this reason, Jackson claims that the sentence is assert*i*ble, but not assert*a*ble. Conversely, if I were paid a large sum of money to make some untrue statement, it may be right to make that statement for pragmatic reasons, in which case, it has some degree of assert*a*bility. However, it would certainly not then be highly assert*i*ble.

The distinction is an important one, and it clearly does some work in Jackson's argument. Assert*i*bility clearly has something to do with the epistemological grounds for asserting a conditional. Here, I judge the assertib*i*lity of a conditional according to the certainty of the belief that an utterance of that conditional would typically produce in the mind of someone who heard that utterance. A conditional is highly assert*i*ble just in case the potential utterer has good grounds overall for thinking this belief will be true. My assertion of the conditional should implicate nothing that I believe is likely to be false. Assert*a*bility, on the other hand, requires that there be good pragmatic grounds for uttering the conditional, and not necessarily that it implicate something I believe to be true. However, we usually have good pragmatic grounds for telling the truth and not asserting anything that might implicate a falsehood, so a conditional's assert*a*bility will often come into line with its assert*i*bility.

There are not many occasions when it is appropriate to assert a conditional when one is certain that the consequent is true. In such cases it is usually more helpful to assert the consequent alone. However, there is an exception in the case of conditionals containing expressions like 'even if' and 'if . . . still'. Suppose I believe that Blair will win the next general election in Britain no matter which Conservative candidate runs against him. It would probably be misleading for me to assert 'If Hague runs against Blair, Blair will win'. Asserting this conditional might suggest that there is some connection between Hague running and Blair winning, presumably to the effect that

Hague is an unfortunate choice of candidate. Now, I might simply choose to say 'Blair will win' and this would certainly be highly assertible. However, I could also legitimately say 'If Hague runs against Blair, Blair will *still* win' or 'Blair will win even if Hague runs against him'. Indeed, my interlocutor would probably find my assertion of either of these conditionals more informative than a straight assertion that Blair will win. This phenomenon requires some explanation.

Like Jackson, I do not believe that it is always right to 'assert the stronger'. If S_1 is logically stronger than S_2 and nearly as highly probable as S_2 (if S_1 is logically stronger than S_2, there will be very few occasions on which it will be equally probable), it is usually right to assert S_1, *but not always*. Therefore, even though I may think that B is true, I may be justified in asserting $(A \rightarrow B)$ rather than B. In particular, I may want to indicate that B will still be true if it should turn out, against the odds, that some other sentence is true that would usually imply the falsity of B (that sentence may be 'A'). In other words, I may want to indicate the *robustness* of B with respect to one or more other sentences.

Consider the following situation. I am watching the World Cup final between France and Brazil. The game is in injury time and France is leading by three goals. Annoyingly, the telephone rings and I answer it, missing the last few minutes of the match. After the telephone call ends, I am certain enough that France were the eventual winner that I do not bother to check the score. Later, somebody asks me who won the match. Clearly, 'France won' is highly probable and would be legitimate for me to assert. However, at the same time, it would surely be equally legitimate, and maybe even preferable, for me to assert 'If Brazil didn't score three goals in injury time, France won' despite the fact that this conditional is logically weaker than ‘France won’ and only marginally more probable. The reason why I would assert the conditional is that it is generally more robust than the non-conditional assertion. In particular, the conditional is robust with respect to ‘Brazil scored three goals in injury time’ while the non-conditional assertion is not. Similarly, suppose a friend tells me later that, although he did not know how many goals France scored, he heard that Brazil scored twice in injury time. I reply ‘If Brazil scored twice in injury time, France won’. Again, my assertion seems legitimate, although it is logically weaker than ‘France won’ and only marginally more probable. If I had just said ‘France won’, I would not be indicating that my assertion was robust with respect to ‘Brazil scored twice in injury time’. Asserting the conditional is more informative, and for that reason often to be preferred.

In the Hague-Blair example, something similar seems to be happening. I assert ‘Blair will win even if Hague runs against him’ rather than just ‘Blair will win’ to indicate that my confidence that Blair will win under the eventuality that Hague runs. If I had just asserted ‘Blair will win’, my interlocutor may not realize that I have taken the possibility of Hague running into account. However, the question of the function of ‘even’ in this assertion remains. Jackson suggests that the function of ‘even’ and ‘still’ is to suggest that the conditional statement containing them is less assertible than a range of related statements of a similar form (Jackson, 1989, p. 44 – 47). The range should be apparent from the position of ‘even’ or ‘still’ together with the conversational context. This phenomenon is easily observed in non-conditional cases. According to Jackson, an assertion of ‘Even Attila can solve the problem’ suggests that ‘Attila can solve the problem’ is less highly assertible than other highly assertible statements of the form ‘ . . .

can solve the problem'. For instance, we might be more confident asserting 'Genghis can solve the problem' or 'Caligula can solve the problem,' even though we think 'Attila can solve the problem' is likely enough true to warrant its assertion. The implication is most likely that Attila is usually not very good at solving problems, and that he can solve this one is surprising. On the other hand, 'Attila can solve even this problem' highlights a range of statements of the form 'Attila can solve . . .' and suggests that 'Attila can solve that problem' is less assertible than other highly assertible statements of this form. It sounds kinder than the other assertion about Attila because it implicates that he can solve a range of problems. These considerations can be applied to conditional sentences containing 'even if' and 'if . . . still'. The use of 'even if' in an assertion of 'Blair will win even if Hague runs against him' suggests that 'Blair will win if Hague runs against him,' though highly assertible, is less assertible than, for instance, 'Blair will win if Clarke runs against him'.

Unfortunately, it does not appear that the robustness of a conditional always determines its assertibility. Consider the following situation. An English teacher is talking to one of her brightest students who is worried about whether he will pass the course. The student is worrying needlessly. His previous work has been so exceptional that he will pass the course no matter how badly he does on the final paper. The teacher, however, does not want to encourage the student to be lazy and says to him, 'If you get an "A" on the final, you will pass the course'. The student then gets a 'B' on the final and duly passes the course. However, he is a little upset afterwards. He asks the teacher why she mislead him that he needed an 'A' when a lower grade would have done just as well. The teacher says, 'I didn't tell you that you wouldn't pass if you got a "B". I just said that if you got an “A” on the final, you would pass. I didn't lie to you!’ Has the teacher mislead the student?

Clearly, the teacher has misled the student in some way. She has asserted the conditional ‘If (A) you get an A on the final, then (B) you will pass the course’purely on the grounds that she knows the consequent is true. However, according to Jackson, this conditional still has a high degree of assertibility. The probability of $(A \supset B/A)$ is high because the teacher knows that B is true however badly the student performs on the final. Therefore, the conditional is highly probable and robust with respect to its antecedent. In fact, this will be the case for all conditionals of which it is known that the consequent is true. For instance, all conditionals whose consequent is a necessary truth will be robust with respect to their antecedents, e.g., ‘If dandelions are delicious, then all pentagons are polygons’. According to Jackson, they will therefore all be highly assertible because, for him, the robustness condition is the *only* thing that needs to be considered when assessing the assertibility of a conditional. Has something gone wrong?

It is clear that the rule of ‘asserting the stronger’ does not apply in all cases and we do sometimes assert conditionals when we know their consequents are true (especially ‘even if” and ‘if . . . still’ conditionals). However, in many cases, such as the case with the teacher, it would be wrong (or, at least, unhelpful and misleading) to assert a conditional on these grounds alone. Though Jackson is right not to adopt the rule that one should always ‘assert the stronger’, the problem with Jackson's account is that all conditionals that are known to be true on the basis of the truth of their consequents (and that are robust with respect to their antecedents) are found to be highly assertible, and this does not seem to be the case. Jackson never tackles this problem satisfactorily in his book.

In order to explain the misleading nature of the teacher's conditional, one might resort to the claim that, although the teacher's conditional is highly assert*i*ble, it is not highly assert*a*ble. In other words, the reasons why the teacher should not have asserted the conditional are of a more pragmatic nature than an epistemological one. Yet, surely this is not the case. What the teacher said caused the student to believe that if he did not get an 'A' then he would not pass the course, and that was a *false* belief. It was not simply that the teacher's remark was unjustified for similar pragmatic reasons as, for instance, a loud utterance of a trivially true statement in a quiet reading room.

A better strategy is to claim instead that high probability and robustness are necessary, but not sufficient, for the assertibility of a conditional. Other contextual factors might be involved. Would an utterance of the conditional imply that there is a causal connection between the antecedent and consequent? Would an utterance of a conditional (A → B) carry with it the implication that (-A → -B) is also true? In many cases, it would, but in other cases, it would not. That the student thought the teacher's conditional carried this implication caused him to be misled. These considerations suggest that a comprehensive and systematic account of the assertibility of conditionals, if such an account is indeed possible, will be highly complex. There is little reason to think that we should be able to calculate assertability by means of a simple probability formula.

I still maintain, however, that robustness is the single most important factor that contributes to the assertibility of a conditional. Therefore, my disagreement with Jackson about the conditions of assertibility for conditionals is perhaps not hugely significant. A more important point of difference between Jackson and me concerns the relation between assertibility conditions and meaning. Jackson claims that the assertibility conditions for a conditional specify its meaning while I do not. On the Davidsonian approach to a theory of meaning that I am following, the meaning of any sentence expressing a proposition is specified by its truth conditions in an appropriate truth theory. If, as Jackson suggests, the conditions governing whether or not it is warranted or justified to assert a conditional determine its meaning, then the Davidsonian proposal must be wrong. Jackson does not give any sustained argument that assertibility conditions, and not truth conditions, establish the meaning of an expression (at least, not when it comes to propositional meaning). He takes it simply as a matter of course that "there is more to meaning than truth conditions" (Jackson, 1991, p. 36). I agree that there is more to meaning than truth conditions when it comes to interrogatives and imperatives. We can also draw a distinction between meaning and truth conditions when it comes to individual expressions that compose sentences. For instance, the expressions 'but' and 'and' are identical in the way that they contribute to the truth or falsity of any expression that contains them. Therefore, as far as truth conditions go, they are equivalent. However, there are several contexts in which 'and' is assertible and 'but' is not, and vice versa. Knowing which connective to use in any given sentence displays one's mastery the English language. If we are to specify the meaning of these expressions purely based on truth conditions, it follows that 'and' and 'but' are synonymous. However, it seems plausible to say that a proper understanding of when to use 'but' and when to use 'and' depends on an adequate understanding of the *meaning* of these expressions, in which case we are forced to say that the meaning of 'but' is different from the meaning of 'and'.

As I will show in Chapter 7, the Davidsonian "truth conditional" approach to semantics has a great deal of utility when it comes to dealing with so-called subjunctive conditionals (though I will not try to defend this approach to semantics here). However, such truth conditional semantics requires us to say that the meaning of a conditional (that is, an assertoric conditional, rather than a conditional command or a conditional question) is specified by its truth conditions. Therefore, I part company with Jackson when he equates the assertability conditions of conditionals and their meaning. It may be, however, that the distance between Jackson and myself is not as great as this suggests. For it may be that Jackson has in mind a less robust conception of truth conditions than I do. For on my view, a sentence's truth conditions in a truth-theoretic semantics, relative to an occasion of utterance, can be given only by a sentence in the metalanguage which expresses the same proposition as that sentence on the occasion of its utterance. Jackson may well have in mind a less stringent conception of truth conditions.

I think that an analysis of indicatives based on robustness works very well given the qualification that robustness is necessary, but not always sufficient for a conditional's assertibility. Curiously though, when it comes to subjunctive conditionals, Jackson supports a traditional Lewis-type possible worlds analysis. I see no obvious reason in principle why we cannot use pragmatic considerations in order to explain our reaction to certain subjunctive conditionals in the same way that Jackson does for indicative conditionals. I take this up in chapter 8, after I have explained my positive proposal in the next chapter.

Some Objections Answered

Grice, Jackson and myself seek to explain the oddness of the assertion of a conditional like 'If Elvis is alive, then he's living in Belgium' on pragmatic grounds. Though we claim the conditional is true (because its antecedent is false), we also claim it would be inappropriate to assert in ordinary conversation because its consequent is not robust with respect to its antecedent. Edgington (1995) argues that this strategy cannot work because it cannot explain why we do not *believe* conditionals like these.

> [T]he difficulties with the truth-functional account cannot be explained away in terms of what is an inappropriate conversational remark [i.e., in terms of assertability conditions]. They arise at the level of belief. Believing that John is in the bar does not make it logically impermissible to disbelieve "if he's not in the bar he's in the library. (p. 245)

Edgington thinks that truth-functional accounts like Grice's and Jackson's fail because

> [w]e need to be able to discriminate believable from unbelievable conditionals whose antecedent we think false. The truth-functional account does not allow us to do this. (p. 245)

The assumption here is that, on the neo-Gricean account, believing that 'John is in the bar' is true makes it logically impermissible to believe that 'if he's not in the bar he's in the library' is false because the truth of the former entails the truth of the latter. It is also true that, if you conducted a survey among non-philosophers, most of them would probably say that they would disbelieve 'if he's not in the bar he's in the library' under such circumstances.

One response that a Gricean could give to such an objection involves a kind of error theory which response I believe Edgington thinks the Gricean must give. Such a response draws a distinction between what people are inclined to believe, and what they ought to believe if they were being completely rational. Using this distinction, the neo-Gricean is able to explain why people might disbelieve ‘if he's not in the bar he's in the library’ though its truth is entailed by their other beliefs. These people are simply in error, though it is an understandable error. In an ordinary context, an assertion of ‘if Bob's not in the bar he's in the library’ implicates the speaker has grounds for believing that there is some connection between its antecedent and consequent. For example, the speaker might have spoken to Bob in the bar just ten minutes ago and believes Bob will head straight to the library after finishing his drink. However, if I know Bob is an illiterate drunkard who spends all day in the bar and never goes near a library, though I can assert the conditional truly, it would be inappropriate to do so.

According to the neo-Gricean, people often say that they disbelieve a conditional because they react to the standard conversational implicature of an assertion of that conditional, not to its meaning. If nobody utters a conditional like the one about Bob unless there is an appropriate connection between the antecedent and consequent, one might be forgiven for believing that the conditional is in fact false if there is no such connection. For ordinary purposes, such an error seems perfectly harmless, and it might even be helpful. If I believe the conditional to be false when no appropriate connection obtains, I am less likely to assert something misleading.[3] Of course, the use of such an “error theory” implies that head-counting evidence is mostly useless for ascertaining the truth conditions of conditionals. However, this does not tell against the neo-Gricean treatment. First, the treatment provides a plausible explanation of why people react to conditionals in the way they do. Moreover, the explanation is in no way *ad hoc* with respect to the rest of the analysis. Second, natural languages are complex structures governed by rules which we do not learn explicitly but rather by catching onto the practice, and most ordinary people are not linguists or semanticists. Mastery of a language in no way requires that one be to explain in any systematic way how its grammar or semantics works. Though any credible linguistic theory should be able to explain why people react to certain utterances in the way they do, we should not expect ordinary people to understand the workings and mechanics of natural languages. So, if ninety percent of non-philosophers react to an assertion of a conditional by disbelieving it, we should certainly be able to explain their reaction, but their reaction itself does not provide good evidence that the conditional really is false.

A more subtle (and, I think, better) way of responding to Edgington's objection was suggested to me by Kirk Ludwig. This response involves specifying what is standardly implicated by an utterance of a sentence of the form 'x believes y'. In saying 'I don't think that if x is in the bar, he's in the library', someone is asserting something false to implicate something true. The general form of what is implicated can be displayed as follows. Suppose that $I_{[s,t]}A$ is the standard implicature of an utterance of A relative to a speaker s at time t. Then we can analyze what is standardly implicated by a sentence of the form 'x believes y' as in (3):

---

[3] To be fair to Edgington, she is perfectly aware of this standard response to her objection. Hers is really a "burden of explanation" argument. She believes that her own theory is preferable on the grounds that it can explain the phenomena without resorting to claims that many of our beliefs about conditionals are in error.

(3) For all individuals x, $I_{[s,t]}$ (x believes A) $=_{df.}$ x believes $I_{[s,t]}$ (A)

According to this analysis, what is standardly implicated by an utterance of the sentence,

(4) 'x believes that if John is not in the bar in the bar, he's in the library'

is that x believes what is standardly implicated by an utterance of the sentence 'if John is not in the bar, then he's in the library'. So, it would not be appropriate to utter (4) if x merely believed that John was in the bar and had absolutely no reason to think that he might be in the library even though the conditional 'If John is not in the bar, he's in the library' is true if John's in the library, as x believes. This explains why people say what they do without convicting them of having any false beliefs at all, and effectively neutralizes this objection to the neo-Gricean account. Why should we think though that implicatures of assertions of the form '*x* believes *y*' are governed by (3)? Well, it would be astonishing if they were not! For when we assert such conditionals, standardly the point of the utterance is to be found in the implicature, and the content of the belief which someone expresses by using such a sentence is given by the content of the implicature. To state explicitly what that belief is there is no better, or more natural, way than to rely on that very sentence the implicature of which when used alone in the context gives the belief's content in the complement clause of the belief report. Consequently, of one announces a belief using a sentence which has a standard conversational implicature, one will be taken to be announcing a belief whose content is given by the content of the standard implicature. Consider a non-conditional case. Suppose that Albert brushed his teeth and went to bed. I would not ordinarily announce this by saying 'Albert went to bed and brushed his teeth', because that would standardly implicate that he did it in the order in which I narrate what he did. For the same reason, however, I would not assert 'I believe that Albert went to bed and brushed his teeth' because that would implicate that I believed that he did it in the order in which those events are expressed in the sentence in the complement clause. Thus, (3) can be justified in general grounds, and supported by cases independent of those at issue here, viz., conditionals, and so its introduction and application to these cases is clearly no ad hoc, but the extension of a general principle governing the relation of standard implicatures of sentences to the standard implicatures of their embedding in attitude reports (and indeed reports of speech).

Another of Edgington's arguments concerns conditional commands. It is plausible, she claims, that if-clauses as play the same role in conditional commands as they do in assertoric conditionals. For example, the if-clause 'If he phones' should play the same role in 'If he phones, Mary will be pleased' and 'If he phones, hang up immediately'. Edgington and I agree that, where possible, one should preserve "semantic innocence" by treating an expression similarly when it occurs in different contexts. One should only admit that an expression is radically ambiguous when absolutely forced to do so.

Edgington claims that preserving semantic innocence in conditional commands leads to difficulties with truth-functional theories like Jackson's. Though Jackson does not extend his theory to account for conditional commands, Edgington speculates that it would go as follows:

> [T]o command that if A, B, is not only to command ($A \supset B$), but also to signal that you would still command ($A \supset B$) if you believed that A were true. (p. 301)

However, Edgington argues that this quickly leads to problems when confronted with examples such as 'If the patient is alive in the morning, change the drip' (as a doctor might say to a nurse). Edgington argues that, on the extension of Jackson's truth-functional analysis she has envisioned, the doctor commands 'Make it the case that either the patient is not alive in the morning, or you change the drip'. Undoubtedly, this claim is based on the logical equivalence of '$(A \supset B)$' and '$(\sim A \vee B)$'. Of course, you could obey this command by simply killing the patient, but that is clearly not what was envisioned when the original command was uttered. Therefore, this truth-functional analysis of conditional commands does not work, and the truth-functional theorist is unable to preserve semantic innocence.

Though I am sympathetic with Edgington's desire to preserve semantic innocence, she is uncharitable to the truth-functional theorist. First, it is far from clear why the truth-functional theories should be forced into the position that Edgington envisages regarding conditional commands, and if she is not forced into that position, Edgington's objection has no force at all. I do not intend to give a full analysis of conditional commands here, but I do wish to demonstrate that there is more than one route the truth-functional theorist could take.

Since commands (whether conditional or not) are neither true nor false, they do not have truth conditions. I therefore give the semantics of these expressions in terms of compliance conditions (based on Ludwig (1997)). An utterance of an imperative sentence is complied with just in case the directive issued by that sentence is obeyed for a comprehensive treatment of non-assertoric sentences in a truth-conditional semantics). On Edgington's envisioned account, this would yield an analysis such as (5):

> (5) For all speakers s, and all times t, a conditional command 'if A, then B' is complied with$_{[s, t]}$ in English iff the person addressed by s at t makes it the case that $(A \supset B)$.

(As I argued in Chapter 5, we must relativize the command to a speaker and a time.) Edgington claims that a problem arises because (5) is logically equivalent to (6),

> (6) For all speakers s, and all times t, a conditional command 'if A, then B' is complied with$_{[s, t]}$ in English iff the person addressed by s at t makes it the case that $(\sim A \vee B)$.

in which case the conditional command could be obeyed simply by making it the case that ~A. However, there does not appear to be any principled reason why the truth functional theorist must give an analysis of conditional commands in the way that Edgington envisions. Moreover, such an analysis is beset with problems from the start.

A conditional command like 'If John phones, hang up' is clearly not analyzable as a command to make a material conditional true. Such an analysis would mean that the conditional is complied with if the person addressed disconnects John's telephone, and that was clearly not what was intended. An analysis such as (7) is a great deal more plausible, and is perfectly consistent with a truth functional account motivated by a desire to preserve semantic innocence:

> (7) For all speakers s, and all times t, 'If John phones, hang up' is complied with$_{[s, t]}$ iff if 'John phones' is true$_{[s, t]}$ then 'hang up' is complied with$_{[s, t]}$.

(8) offers a general account of conditional commands along these line:

> (8) For all speakers s, and all times t, a conditional command 'If A, then B' is complied with$_{[s,\,t]}$ iff if A is true$_{[s,\,t]}$ then B is complied with$_{[s,\,t]}$.

According to (8), if A is false, the person addressed by the conditional command is not obliged to do anything to obey that command, and this certainly seems to be what is intended when such a command is uttered. Of course, S could see to it himself that A is false, but he has in no way been commanded to do so. Moreover, in the utterance of a conditional command, there might well be an implicit understanding that the person addressed should not see to it that antecedent is false (as seems likely in the patient case). Therefore, Edgington's objection is easily dispelled by a more plausible account of conditional commands (similar remarks apply to conditional questions).

.

Summary

In this chapter, I have sought to establish the plausibility of an analysis of indicative conditionals based on the thesis that indicative and material conditionals are logically and semantically equivalent. This, I believe, is the simplest and most elegant theory of indicative conditionals. It explains the raw data about how people react to the assertion of various conditionals without forcing use to abandon the truth functional analysis of the familiar logical connectives. Moreover, it explains how there can be a point to uttering a conditional that is trivially true. I do not claim that no other theory of indicative conditionals is consistent with the theory of so-called counterfactual conditionals that I offer in the next chapter. However, any such theory must explain how it can sometimes be appropriate to utter a trivially true conditional, or even a false one. I do not claim to have mounted a comprehensive defense of this theory of indicatives (that would be a book-length project by itself), though I have defended it from some salient objections. I believe that this treatment of indicatives provides the foundation for a most elegant analysis of so-called counterfactual conditionals. It is to this analysis that I now turn.

CHAPTER 7
A TRUTH-THEORETIC SEMANTICS FOR CONDITIONALS

Auxiliary verbs such as 'were', 'had', and 'would' appear in so-called subjunctive conditionals where they would not appear in so-called indicative conditionals. As I showed in Chapter 2, in other contexts, we understand these verbs to be part of tense constructions. However, almost everyone who writes about conditionals ignores this. Instead of taking the meaning of these tensed constructions at face value, they argue that these expressions indicate a special class of conditionals called 'subjunctive' or 'counterfactual'. They then construct semantics for this special class, usually based upon possible worlds. Whether or not these special constructions indicate that a conditional statement is about possible worlds is certainly not a trivial issue. The introduction of possible worlds into an account of conditionals is likely to have serious ontological implications, and even if one attempts to treat possible worlds in a metaphysically innocuous way, e.g., as maximally consistent sets of sentences, as, for instance, Pollock (1976) does, it is not intuitive that so-called subjunctive conditionals are really about such entities.

However, avoiding an ontology of possible worlds is not the primary motivation behind my account of so-called counterfactual conditionals. The preservation of semantic innocence should be an important consideration in the construction of any semantic theory. Where possible, one should treat the same expression in the same way when it occurs in different contexts. One should only admit that an expression is radically ambiguous between different contexts when absolutely forced to do so. It must therefore be preferable to take the tense of the verbs that occur in conditional constructions at face value if we can. If they look like past-tensed verbs, we should endeavor to treat them as such.

I refer to the conditional clause of a conditional statement as the 'protasis', while I call the main clause the 'apodosis'. I prefer to avoid the expressions 'antecedent' and 'consequent' whenever I can. My reason is that what is usually called the 'antecedent' of a conditional is frequently something other than what is evident from its surface grammar. For example, authors like Adams and Lewis say that the antecedent of 'If Oswald had not killed Kennedy, then somebody else would have' is 'Oswald did not kill Kennedy' rather than 'Oswald had not killed Kennedy'. As I showed in Chapter 2, they change the past perfect verb construction in the protasis of the original conditional ('had not killed') to a construction of the simple past ('did not kill'). I believe that this reconstruction of antecedents is the cause of much confusion.

Most of the conditionals that philosophers call 'subjunctive' or 'counterfactual' involve what linguistics call a 'backshifting' of tenses in the conditional clause. According to the orthodox view, a 'backshifted' conditional is one in which the tense of the verb in the conditional clause is shifted to some tense usually used to pick out a time further into the past than the time of the hypothetical event which the clause is used to talk about, e.g., the present is often shifted to the simple past, the simple past is often shifted to the past perfect. Consider the following examples:

(1) If I were president, I would improve the public school system.
(2) If I were elected president in 2000, I would improve the public school system.

(3) If I had seen you, I would have waved.

The conditional clause of these 'backshifted' conditionals is standardly implicated by the speaker as expressing a hypothesis that is contrary to fact. According to Dahl (1997), the received view of the semantics of these conditionals among linguistic theorists is that the past tense in cases like (1) and (2) is used for present and future time reference and that the past perfect in cases like (3) is used for past time reference. This is accompanied by the view that the backshifting of the tense conveys a belief on behalf of the speaker about the fulfillment of the condition described in the conditional clause, e.g., that it is contrary to expectation, contrary to assumption or contrary to fact.

I deny that conditionals like (1), (2) and (3) literally *mean* something about the present or future and argue that the tense of the verb in the conditional clause is about exactly that time, relative to the time of the utterance, that one would expect it to be about. Any suggestion of a time other than that indicated by the tense of the verb is the result of the expected conversational implicature of an utterance of the sentence in which that verb occurs. Therefore, there is nothing particularly special about the semantics of backshifted conditionals. The only significant *semantic* difference between these conditionals and most ordinary indicatives is that they contain past-tensed verbs. For instance, the protasis of (1) contains the expression 'I were president', a more stylized equivalent of 'I was president', which is the past tense of 'I am president'. The protasis of (3) contains the expression 'I had seen you' which is the past perfect tense of 'I see you'. Any other difference between these conditionals and others is entirely *pragmatic.*

Table 1 lists some examples of conditionals classified by the tenses of the verbs that appear in their protases and apodoses. The table shows some conditionals with past tense verbs standardly used to talk about hypothetical events at different times. Of course, there are many other possible combinations of tenses and uses other than those listed here, but the combinations in the table are the ones that occur in so-called counterfactual conditionals. The table also contains two types of conditional, which are not standardly called counterfactual (the first two in the list) but which are often discussed in the literature on so-called counterfactuals.

Table 3: Examples of past tense conditionals.

Verb in protasis	Verb in apodosis	Example	Example standardly used to talk about
Simple past	Simple past	If Oswald didn't kill Kennedy, someone else did [kill Kennedy].	the past
Simple past	Future-in-the-past	If Nixon ate steak, he would drink a glass of wine.	the past
		If the Germans invaded, they would win the war.	the present
		If I ran for President in 2000, I would lose.	the future
Past perfect	Future-perfect-in the-past	If Oswald hadn't killed Kennedy, someone else would have [killed Kennedy].	the past
		If Jim had been here, he would have complained.	the present
		If Jim and Sue had married, they would have been miserable. (uttered just after the wedding has been called off)	the future

I will refer to those conditionals that have the simple past or the past perfect in their protases and a relativized future modal (the future-in-the-past or the future-perfect-in-the-past) in their apodoses as 'backshifted'. This results in a class of backshifted conditionals slightly broader than the one traditionally defined. This class of conditionals includes all those that philosophers usually refer to as 'subjunctive' or counterfactual', as well as a few others. My goal here is to provide an analysis of the semantics and conversational role of backshifted conditionals. Since the class of backshifted conditionals includes those conditionals usually described as ‘counterfactual’, by providing a semantics for this class of conditionals, I solve the traditional problem of providing a semantics for ‘subjunctive’ or ‘counterfactual’ conditionals.

A semantics for backshifted conditionals

Conditionals of the simple past/future-in-the-past combination used to talk about the past

I will illustrate my proposal for a semantics of backshifted conditionals by providing a semantic analysis of each of the types of conditional in table 1. I hope the relative simplicity and elegance of my proposal will convince the reader of its correctness.

Let us first look at conditionals with the simple past in their protases and the future in the past in their apodoses. Consider the following:

(9) If Nixon ate steak, he would drink a glass of wine.

This conditional appears to say something about certain actions that Nixon performed habitually. It is similar in import to ‘Whenever Nixon ate steak, he would drink a glass of wine’. Without introducing a great deal of theoretical baggage, we can recognize the following five elements of a conditional sentence like (9):

The speaker time t*, i.e., the time at which the sentence is uttered
The time t_p of the hypothetical event described in the protasis (Nixon eating steak)
The time t_a of the hypothetical event described in the apodosis (Nixon drinking a glass of wine)
The tense of the verb in the protasis (‘ate’), i.e., the past tense
The tense of the verb in the apodosis (‘would drink’), i.e., the future in the past

My proposal is that the tense of the verb in the protasis, i.e., the simple past, locates t_p relative to t* just as the simple past locates the time of Nixon eating steak relative to the speaker time in the following non-conditional sentence:

(10) Nixon ate steak.

In a similar fashion, I suggest that the tense of the verb in the apodosis of (9) locates t_a relative to t_p, though here the tense is the future in the past.

We can capture the meaning of (9) with the following semantic treatment:

(9a) [There is a t_1: $t_1 < t^*$](if Nixon eats$_{t1}$ steak, then [there is a t_2: $t_2 > t_1$](Nixon drinks$_{t2}$ a glass of wine))

This analyzes the conditional construction in the simplest possible way that is consistent with our analysis of tensed verbs in non-conditional constructions.

One reason that this treatment might seem odd is that it involves an existential quantifier, when the natural reading of (9) might seem to require a universal quantifier. My reason for treating (9) as (9a) is motivated by a suggestion of Kirk Ludwig's. The suggestion is to treat 'donkey' conditionals with anaphoric pronouns such as (11)

(11) If a man owns a donkey, he beats it

as having the form of (11a)

(11a) [Some x: x is a man][some y: y is a donkey] (if x owns y, then x beats y)

rather than as (11b)

(11b) [For all x: x is a man][for all y: y is a donkey] (if x owns y, then x beats y)

It appears that the ‘he’ and ‘it’ in (11) are pronouns functioning as variables bound by ‘a man’ and ‘a donkey’ respectively, in which case they both would have wide scope, as in (11a) and (11b). Contrary to what I have supposed, the natural reading of (11) might seem to be the reading captured with the universal quantifier in (11b) rather than the existential quantifier in (11a). One response to these phenomena consistent with this intuition is to claim that constructions of the form

Indefinite article + noun,

such as ‘a man’ and ‘a donkey’, are ambiguous when used as noun phrases between a universal and an existential reading. However, in many cases similar to (11), we do not get the universal reading. Consider the following sentence:

(12) A man owns a donkey and he beats it.

The only difference between (12) and (11) is the connective. In one case it is ‘if . . . then’ while in the other it is ‘and’. Yet (12) clearly requires a treatment with a particular quantifier, i.e.,

(12a) [Some x: x is a man][some y: y is a donkey] (x owns y and x beats y)

A treatment with a universal quantifier does not seem to be available here. This makes the claim that there really is an ambiguity in these cases appear suspect and motivates the claim that (11a) is in fact the correct semantic form for (11). We can then derive a 'universal' reading for this conditional as the result of certain expected pragmatic implications of the semantic treatment in (11a). The import of an utterance of (11a) can be shown to be the same as the import of an utterance of (11) when the standard import of uttering (11) is treated as a conversational implicature.

A possible response here is that there are cases similar to (12) where a treatment with a universal quantifier seems appropriate. For instance, 'A man is born and he dies' might seem to have the form, '[For all x: x is a man](x is born and x dies)' rather than '[There is an x: x is a man](x is born and x dies)'. Yet, there are good reasons for not treating the sentence with the universal quantifier.

First, consider the use of the present tense in 'A man is born and he dies'. The use is a peculiar one, since anyone who uttered such a sentence would not mean to say that a man is born (or men in general are born) in the present and die in the present. This becomes more obvious when one adds an adverbial, such as in 'A man is born, and he dies seventy years later'. These considerations suggest that the import of an assertion of 'A man is born and he dies', whether it is treated with a universal or particular quantifier, cannot be explained without referring to pragmatic considerations; otherwise, we simply cannot make sense of it. This allays one motivation a person might have for using the universal quantifier, i.e., that treating it with a universal quantifier does not require any pragmatic explanation.

Second, consider 'A man is born and he dies' when its present tense verbs are shifted into the future or the past:

A man will be born and he will die.
A man was born and he died.

In both of these cases, a 'universal' treatment does not seem to be available. It would be puzzling if shifting the tense of the verbs of a sentence should make a difference to the quantifier. This suggests that we ought to treat the sentence with a particular quantifier if we can.[1]

One might also be tempted to argue that (11b) is the more plausible treatment because treating (11) as (11b) does not require one to derive the standard interpretation of an assertion of (11) using pragmatic considerations. The standard interpretation, one might say, is laid bare by the semantic treatment. However, this response is misguided. Even if one were to treat (11) as (11b), one would still need to provide an explanation of how one derives the standard conversational import of an utterance of (11) on the assumption that (11b) is the right semantic treatment. (11b) is true if nobody owns a donkey (given our assumption that the indicative conditional is semantically equivalent to the material conditional), but no interlocutor would ever take this possibility into account when considering the import of someone's utterance of (11), because nobody would assert (11) in such circumstances. If one analyzes (11) as (11b), one is obliged to explain this phenomena. Therefore, whichever analysis one chooses, a pragmatic explanation of the conditional's assertability conditions is needed. The fact that analysis (11a) requires such an explanation does not make it any less plausible than (11b), and given the considerations mentioned above, the weight of evidence favors treating the conditional as (11a).

---

[1] Later in this chapter, I will give some examples of how, in certain situations, pragmatic considerations can lend a "universal" reading to a particular quantifier.

On the assumption that (11a) is the right semantic treatment of (11), one would not assert (11) unless one thought there was some special (non truth-functional) connection between a man owning a donkey and him beating it. The reason is that (11a) is trivially true, since its truth merely requires that there is a man and a donkey such that the first does not own the second. (Also, it is true merely if some man beats some donkey, though this is not quite as trivial as some man not owning some donkey.) The only point to asserting (11a), then, would have to derive from the speaker's having some ground for thinking it true which did not depend on her knowing that some (or every) pair of a man and a donkey failed to satisfy the open sentence 'x owns y' or satisfied the open sentence 'x beats y'. Otherwise, she would be violating Grice's maxim of quantity: she would be being less informative than she could be in the circumstances. She could have asserted either that some man owns a donkey or that some man beats a donkey or the conjunction of the two. By choosing instead to utter (11), the speaker (on the assumption she is obeying Grice's Cooperative Principle), licenses her interlocutor to suppose that she is not asserting (11) simply on the basis of knowing some (or every) man donkey pair fails to satisfy 'x owns y' or satisfies 'x beats y'. This requires her to know that a man-donkey pair satisfying 'x owns y' also satisfies 'x beats y', whatever man-donkey pair you take, in virtue of some connection between the relation of owning holding between a man and a donkey and the relation of beating holding between them. And this is precisely the implicature which is predicated by Grice for the universally quantified 'For every man, and every donkey, if the first owns the second, he beats it'. This explains then why we generally take a sentence like 'If a man owns a donkey, he beats it' to have a universal reading, i.e., to be the same in import as a corresponding universally quantified sentence.

Moreover, in deriving the import of an assertion of this conditional, we have not taken any controversial philosophical move other than what would be sanctioned by any standard Gricean analysis of conversational pragmatics.

One would therefore not assert (11) unless one wanted to imply some special connection between the relation expressed in the antecedent and that expressed in the consequent as holding between man-donkey pairs. This implication is exactly what one gets on the assumption that (11a) is the right semantic treatment for (11). Together with the evidence gathered from comparing (11) with similar cases like (12), we now have a compelling reason to believe that (11a) is the right semantic treatment for (11). By treating the standard interpretations of these conditionals as the result of the expected pragmatic implications one obtains from the proposed semantic interpretation, we can derive the universal interpretation of (11) that we want by treating it as (11a).

We can achieve the same effect by treating (9) as (9a).

(9a) [There is a t_1: $t_1 < t^*$](if Nixon eats$_{t1}$ steak, then [there is a t_2: $t_2 > t_1$](Nixon drinks$_{t2}$ a glass of wine))

If someone asserts something with the semantic form of (9a), we must suppose that some time *x*, which is earlier than t*, satisfies (9b):

(9b) If Nixon eats$_x$ steak, then [there is a t_2: $t_2 > x$] (Nixon drinks$_{t2}$ a glass of wine)

(9b) is trivially true, since there will be many past times at which Nixon did not eat steak. It would not be conversationally helpful to assert (9) on the grounds that it is trivially true. There must be some other non-trivial (i.e., non-truth functional) reason for the speaker to assert it. This reason must surely be that there is some connection between Nixon eating steak at a time and him drinking wine at a later time such that he always drinks wine whenever he eats steak. In other words, by treating (9) as (9a) we get the 'universal' reading we want, just as we did by treating (11) as (11a). This suggests that we treat both (9) and (11) using existential quantifiers if we can.[2]

Backshifted conditionals used to talk about the present or future

We have analyzed conditionals of the simple past/future-in-the-past combination when the past tenses in the protasis and apodosis are used to talk about habitual or recurring events in the past, but how do we deal with backshifted conditionals like (13) when they are used to talk about the present or future?

(13) If the Germans invaded, they would win the war.[3]

It is commonly thought among linguistic theorists that past tenses in languages like English are sometimes used without past time reference, and this is particularly evident in the case of backshifted conditionals. (13) might seem to be just such a case, since the past tensed 'missed' looks like it is being used to talk about a hypothetical event in the present or the future. It is not difficult to find examples of the simple past, as well as the past perfect, being used in this way. Consider the following examples:

---

[2] Sometimes conditionals like (9) are used to say something about events that are frequently connected rather than invariably connected. So the import of an assertion of (9) may not be that Nixon always drunk wine whenever he ate steak, but that he did so frequently. A plausible explanation of this phenomenon consistent with my analysis is that, when one makes such an assertion, one's interlocutor reads in an adverb, such as 'typically', that is not stated explicitly. So, one might understand an assertion of (9) to have the same meaning as 'Typically, if Nixon ate steak, he would drink a glass of wine.' Though I will not discuss the treatment of adverbials here, treatment (9a) easily allows us to treat the sentence with the adverb. Construing 'typically' as a sentential operator, we obtain the following.

> It is typical of Nixon that [there is a t_1: $t_1<t^*$](if Nixon eats$_{t1}$ steak, then [there is a $t_2>t_1$](Nixon drinks$_{t2}$ a glass of wine)).

[3] This is equivalent in meaning to 'If Germany were to invade England, they would win the war.'

(14) If I had enough money, I would go and visit him.
(15) If he had been alive next year, he would have been 200 years old.
(16) If war had broken out tomorrow, I would have enlisted.[4]

Dahl (1997, pp. 103 – 107) also gives a number of similar examples where the past perfect is apparently used for present or future time reference from other European languages such as Bulgarian, German, and Swedish.

One might argue that these cases are just aberrations, i.e., that they are not properly assertable, but there are enough examples of them to make this view implausible. Moreover, it is difficult to see how one could express the message conveyed by (16) in any way that does not use the past perfect. For instance, (17),

(17) If war were to break out tomorrow, I would enlist.

does not seem to carry with it the implication of (16) that any threat of war has now passed. Therefore, it is not plausible to say that (16) is merely a variant of (17).

A better way to treat these cases is to take them at face value, i.e., to treat the past tense verbs as having past reference. We can then derive the standard interpretation of these conditionals by examining the pragmatic implications of their assertion. So, we should treat (13) in a similar way to how we treated (9):

---

[4] We can imagine that someone might say this after a situation in which war had seemed immanent but was averted at the last minute.

(13a) [There is a t_1: $t_1 < t^*$] (if the Germans invade$_{t1}$, then [there is a t_2: $t_2 > t_1$] (the Germans win$_{t2}$ the war))[5]

We can then derive the standard interpretations of (13) from the expected pragmatic implications one would obtain from uttering the conditional based on the analysis in (13a).

Why one would assert (13) (or, indeed, (14), (15), or (16)) if one intended to talk about a (hypothetical) event in the present or future rather than in the past, as the tense seems to suggest? It is usually clear enough from the context, or from the use of adverbials, when one intends to use past tenses to make a conditional claim about the present future, but why use the past time reference at all? The answer, I believe, is that it is the past time reference that gives an assertion of (13) its pragmatic import. Consider the following three sentences:

(13) If the Germans invaded, they would win the war.
(18) If the Germans invade, they will win the war.
(19) If the Germans had invaded tomorrow, they would have won the war.

---

[5] This proposal entails that some sentences involving past time reference applied to the future will turn out to be false but still assertable. According to Grice, you can utter something that is false if doing so serves some conversational point, usually by implicating something which one believes to be true. Grice himself gives an example of a sentence that would be conversationally assertable though false:

(a) It is not the case that if God exists, then we are free.

This is literally equivalent to,

(b) God exists and we are not free.

But someone who asserted (a) would usually not be taken to have committed himself to (b). We merely take the person to be negating the *standard implicature* of an assertion of

(c) If God exists, then we are free,

rather than asserting (b) which is equivalent to the *literal* negation of (c). According to Grice/Ludwig, practices like one this arise in the light of a standard practice grounded in conversational dynamics.

I maintain that (13) and (19), both of which contain past tense verbs, have a certain pragmatic implication that (18) does not have which explains why they are used for talking about hypothetical future events.

First let us consider (19). The import of (19) is best seen by comparing it to a more obviously past-referring use of the past perfect/future-perfect-in-the-past combination. Consider (20):

(20) If the Germans had invaded last week, they would have won the war.

A usual pragmatic implication of an assertion of (20) is that any German invasion subsequent to the time of utterance would not result in Germany winning the war. The point to asserting (20) must derive from the speaker's having some ground for thinking that a future German invasion would not be successful. Otherwise, he would be violating Grice's maxim of quantity by being less informative than is appropriate in the circumstances. If he thought that a future German invasion could still result in victory, it would be less than helpful to simply assert (20). He should say something about the future threat, not just the threat at some time in the past. We have excellent reason to believe this is a *pragmatic* implication because it can be cancelled, as in the following example:

> (21) If the Germans had invaded last week they would have won the war, but they would still win if they invaded tomorrow.

The use of the past perfect (or the 'past in the past') in (20) suggests that, although there may have been a possibility of an invasion resulting in a German victory in the past, that possibility is now over and done with. The factors that might have lead to a German victory have passed into history. Yet this is surely what we also want to imply when we assert (19), i.e., you would not assert (19) if you did not think that any possibility of invasion had vanished (or, if you did, you would need to indicate that you were ‘opting out’ of the usual conversational implicature). So, one uses (19) when one wants to say that Germany will not invade tomorrow but also that conditions are favorable then for German success. Generalizing from this case, I want to claim that we use the past perfect in the protasis of a conditional when we believe that the possibility of the conditional clause actually being fulfilled is a *closed possibility*, i.e., when we believe that it may have been a real possibility in the past, but that it is not any more. This applies whether or not the conditional clause is intended to say something about an event in the past, the future or the present. Apart from allowing us to treat the ‘had invaded’ in (19) as a case of the past perfect (which, after all, is exactly what it looks like), this account of such uses also gives a neat explanation of why we very seldom use the past perfect when talking about future events. We know less about what will happen in the future than we do about what has happened in the past. Crudely speaking, this is because the future just has not happened yet! Therefore, we are less often able to say that the possibility of an event happening in the future is a closed possibility.

I have explained why one would assert (19), but what about the difference between (13) and (18)? (13) is to be treated as (13a)

> (13a) [There is a $t_1{:}t_1 < t^*$] (if the Germans invade$_{t1}$, then [there is a t_2: $t_2 > t_1$] (the Germans win$_{t2}$ the war))

As we saw with (9a), we can use pragmatic considerations to derive a 'universal' reading from this treatment. In the case of (13), it is similar in import to 'whenever the Germans invade (under given circumstances), they win the war.' This, I suggest, is exactly the reason why we use the past tense in conditionals like (13) when talking about the future, i.e., because we want our assertion to have universal import. The reason why one would assert (13) is that one thinks that there is some necessary or lawlike connection between the Germans invading under circumstances similar to those occurring tomorrow and the Germans winning the war. The reason why there is a difference between (18) and (13) in this regard is because, according to my analysis of (13), a quantifier binds across the conditional, but this is not so for (18). This explains why (13) offers the 'universal' interpretation that (18) does not. Moreover, one might still utter (13) if one thought there were *absolutely no possibility* of the Germans invading tomorrow (though, if one really wanted to show there was no possibility, one might then be inclined to use (19) instead). Sentence (13), on the other hand, because it does not have the same universal import, does not share this feature. Therefore, (13) would not be an appropriate assertion if one thought there was no possibility whatsoever of a German invasion tomorrow.

A good indicator that the standard import of utterances of conditionals of this type depends upon pragmatic considerations is that utterances of those conditionals do not have their standard import in certain contexts. In effect, this is similar to showing that utterances of these conditionals have standard implicatures that can be cancelled in the way that Grice suggests. Consider the following sentence:

(22) If I had a million dollars, I would own a yacht.

Under normal circumstances, we would understand the import of an utterance of such a conditional as saying something about a counterfactual situation in which the speaker has a millions dollars. According to my analysis, the meaning of (22) is given by (22a)

> (22a) [There is a t_1: $t_1 < t^*$] (If I have_{t1}, a million dollars then [there is a t_2: $t_2 > t_1$] (I own_{t2} a yacht))

So, (22) is literally saying something about a hypothetical event in the past. Ordinarily, we would not understand a speaker who uttered (22) as being concerned with the past, so we understand him as trying to say something about the present. In this case, he appears to be indicating some sort of connection between him (counterfactually) having a million dollars and him owning a yacht thereafter.

However, it is quite possible to imagine a scenario in which we do not get this import from an utterance of (22). Suppose I am recovering from a bout of amnesia and I am trying to remember my possessions for legal purposes. I have good reason to believe that I had a million dollars at some time in the past and I know that I love sailing. There is a yacht moored in the local harbor, and I suspect it might be mine. So I say, ‘if I had a million dollars (that is, in the past), I would own a yacht (that is, now)’. Thus, the import obtained from an utterance of (22) in a more normal context does not arise here.

An important virtue of my account is that it is able to handle this phenomenon without admitting that (22) is ambiguous (as one would have to do in an account based on the traditional distinction between conditionals). Thus, the account presented here best preserves Grice's "Linguistic Occam's Razor" by not multiplying the meaning of conditional expressions beyond necessity.

Conditionals of the past perfect/future-perfect-in-the-past combination used to talk about the past

I have shown how conditionals with past tense verbs can be used to talk about hypothetical present and future events. However, I have not yet explained the use of conditionals with the past perfect in their protases and the future-perfect-in-the-past in their apodoses nor how it differs from the use of conditionals with the simple past in both their protases and apodoses. We sometimes feel justified in asserting conditionals of one type when we would not assert a conditional of the other type. Much ink has been spilled concerning Ernest Adams' (1970) examples:

> (23) If Oswald didn't kill Kennedy, somebody else did [kill Kennedy].
> (24) If Oswald hadn't killed Kennedy, somebody else would have [killed Kennedy].

Most people would agree that (23) is legitimately assertable, but asserting (24) is highly contentious. This has prompted many to claim that (23) and (24) are fundamentally different kinds of conditionals. I reject this claim. However, to make my position plausible, I need to explain the difference in reaction between (23) and (24).

Brian Ellis has an interesting response to (24). He asks us to imagine that it is November 1963 and Mary believes there is a conspiracy to kill the president. She says:

> (25) If Oswald doesn't kill Kennedy then someone else will. (Ellis, 1984, p. 52)

Knowing now that Oswald killed Kennedy, Ellis claims that we use (24) to say now what Mary said back in 1963. He concludes from this that (24) and (25) are just different ways of saying the same thing. The only difference is that their utterance is appropriate to different times and circumstances. Ellis is a traditionalist in the sense that he buys into the illusionary indicative/subjunctive distinction. However, in another sense, his view is highly unconventional. Since he believes them to be synonymous, he concludes that there is no logical distinction between the 'indicative' (25) and the 'subjunctive' (24). This motivates him to treat both indicatives and subjunctives as non-truth functional. In this respect, he likens himself to Robert Stalnaker, who prescribes a possible worlds analysis for both kinds of conditional. Of course, if he had been thinking more clearly, he would have used these considerations to conclude that the verbs in (24) are past-tensed, not subjunctive and that there really is no distinction of mood here.

Ellis's remarks are not altogether off target, though they need correcting in two ways. First, he is mistaken to suppose that (25) and (24) say the same thing. If we believe that meanings of utterances are relativized to both speakers and times, as seems plausible, they cannot be synonymous (though I suspect Ellis uses the term 'synonymous' in a very loose sense, so I am probably being unfair). However, even if the two utterances are not synonymous, it is certainly right that if (25) is an appropriate thing to say now, then (24) must have been the right thing to say back in 1963.

The second problem is that Ellis does not show enough sensitivity to the tenses of the verbs in (25) and (24). If we shift the present tense verbs in (24) to the simple past we get:

(26) If Oswald didn't kill Kennedy, then someone else would.

This seems to more accurately express now what Mary thought in 1963, and it does not implicate that Oswald actually killed Kennedy, which is something Mary could not have known at the time. Furthermore, (26) is exactly what we would use when indirectly reporting Mary's belief, as in (27):

(27) Mary believed that if Oswald didn't kill Kennedy, then someone else would.

Remembering considerations such as these will help us give a plausible account of (23) and (24) that properly considers their use of tense expressions.

To avoid the complication introduced by the quantifier expression 'somebody', let's pretend that Ruby was the only other person who could have killed Kennedy but that we have no reason to suppose that he would have. We can then consider the following (historically implausible) alterations of the above.

(28) If Oswald didn't kill Kennedy, Ruby did [kill Kennedy].
(29) If Oswald hadn't killed Kennedy, Ruby would have [killed Kennedy].

It's natural to treat 'hadn't killed' in (29) as the past perfect, and to treat 'would have killed' is as the future-perfect-in-the-past. In (28), the verbs in the protasis and the apodosis both seem to be in the simple past.

The negations in (28) and (29) cause additional complications so, for the moment, I will consider the following simpler conditional:

(30) If Oswald had killed Kennedy, he would have fled the scene of the crime.

The tense of the verb in the protasis of (30) locates t_p (the time of the hypothetical event described in the protasis) relative to t* just as that of 'had killed' locates the time of Oswald killing Kennedy relative to speaker time in the non-conditional (31), which receives the treatment (31a).

> (31) Oswald had killed Kennedy
> (31a) [There is a t_1: $t_1 < t^*$] ([There is a t_2: $t_2 < t_1$] (Oswald kills_{t2} Kennedy))

In a similar fashion, I suggest that the tense of the verb in the apodosis of (26) locates t_a (the time of the hypothetical event described in the apodosis) relative to t_p, though here the tense is the future-perfect-in-the-past rather than past perfect.

I propose to treat the tenses that appear in conditional constructions at face value. It is eminently reasonable to treat the protasis of (30) in the same way we treat (31). The protasis of (30) looks exactly like (31), and treating it this way is the simplest analysis consistent with our treatment of tense in non-conditional constructions. So, we should expect (31a) to figure in our semantic treatment of (30). Likewise 'Oswald would have fled . . .' should receive its usual treatment in (30), i.e., as a past-shifted future modal. Thus I represent (30) as (30a).

> (30a) [There is a t_1: $t_1 < t^*$] (if [There is a t_2: $t_2 < t_1$] (Oswald kills_{t2} Kennedy), then
> [There is a t_3: $t_3 > t_1$][There is a t_4: $t_4 < t_3$] (Ruby flees_{t4} . . .))

In (30a) I put the indexed restricted quantifier out front because the past-shifted future modal in the apodosis shows that it is controlled by the tense of another verb in the sentence, which here must be that of the protasis, but this requires that it take wide scope over the conditional. For (29), I propose the analysis (29a), and for (28), (28a).

> (29a)[There is a t_1: $t_1 < t^*$] (if ~[There is a t_2: $t_2 < t_1$] (Oswald kills$_{t2}$ Kennedy), then
> [There is a t_3: $t_3 > t_1$][There is a t_4: $t_4 < t_3$] (Ruby kills$_{t4}$ Kennedy))
> (28a) If [There is a t_1: $t_1 < t^*$] ~(Oswald kills$_{t1}$ Kennedy), then [There is a t_2: $t_2 < t^*$] (Ruby kills$_{t2}$ Kennedy)

To make these treatments plausible, we need to consider the standard pragmatic implications one would get from asserting (29a) and (28a). First, by interpreting (29) as (29a), we can see that someone who asserts (29) will implicate she has non-truth-functional grounds for thinking some past time satisfies (29b) with respect to 't_1'.

> (29b) if ~[there is a t_2: $t_2 < t_1$](Oswald kills$_{t2}$ Kennedy) then [there is a t_3: $t_3 > t_1$][there is a t_4: $t_4 < t_3$](Ruby kills$_{t4}$ Kennedy).

In fact there is a past time t_1 which makes the antecedent false, namely, the time Oswald killed Kennedy. For there is a past time such that there is an earlier time at which Oswald killed Kennedy. So, a speaker can assert (29a) truly on this interpretation of it. However, she does not assert it on the grounds that some time makes the antecedent false, for if she did, she would violate Grice's maxim of quantity. She could have just asserted, 'Oswald killed Kennedy', and be done with it. So, to see her as obeying the Cooperative Principle, we must see her as intending to convey that she has non-truth-functional grounds for thinking some past time satisfies (29b). General background knowledge is sufficient for us to infer that the speaker does not think that any time will do. So we can assume that she intends to implicate that she has non-truth-functional grounds for thinking some time–likely the time of Oswald's killing of Kennedy, since we know she thinks Oswald did kill Kennedy–satisfies (29b). What would such grounds be? Since we know the speaker thinks Oswald did kill Kennedy, we can infer that she thinks that there are grounds for thinking that a process in place would lead to Ruby's killing Kennedy in the event that Oswald did not preempt him. That is, Kennedy's death was over-determined, as someone's death who is poisoned and then shot before the poison takes effect. Depending on the context, a more specific implication about that process might emerge, e.g., that it was a conspiracy and that Ruby was the backup man on the grassy knoll, or that independently Ruby had a grudge against Kennedy and was bent on homicide, but slower in his preparations, etc. In any case, minimally, what is implicated is that but for occurrence of an event of the sort expressed by the main verb in the protasis, the event expressed in the apodosis in due course would eventuate. The usual point of an utterance of (29) is surely to convey just this information. Thus, by considering the pragmatics of an utterance of (29a) its standard interpretation is derived.

The position of the negation in (29a) is crucial to this analysis. What grounds are there for thinking that it should take wide scope with respect to the past tense marker on 'killed'? Consider (32).

(32) They had not finished by midnight.

Intuitively, this says that at no time prior to midnight were they finished. We can give the negation wide or narrow scope with respect to the past tense marker as in (32a) and (32b).

(32a) [There is a $t < t^*$] ($\sim$[there is a $t_1 < t$ & $t_1 <$ midnight of t_1] (they$_{t1}$ finish))
(32b) [There is a $t < t^*$] ([there is a $t_1 < t$ & $t_1 <$ midnight of t_1] $\sim$(they$_{t1}$ finish))

(32a) is the right analysis while (32b) is clearly wrong. (32b) could be true even if they *did* finish before midnight. So we have good non-*ad hoc* grounds for representing the logical form of (29) as (29a).

Turning to the analysis of (28), two things seem salient. First, it is a general feature of conditional constructions which have the past perfect with 'not' in their protases that the event described in the protasis is suggested to be contrary-to-fact. In other words, conditionals which follow the general pattern,

Had X not happened, then Y would have/not have happened.

carry with them the (pragmatic) implication that X *did* happen or is strongly expected to have happened.[6] The event described in the protasis is something we believe happened, but the use of this particular construction indicates that the speaker is considering what would follow under the assumption that it did not happen.[7] This is a feature shared by the past perfect together with 'not' in non-conditional contexts. In such contexts, when we assert sentences involving the past perfect together with 'not', we seem to imply that something *that was or is expected* did not take place or did not take place until after the time referred to. This, I maintain, is a standard conversational implicature of any assertion of these sentences. Consider the following:

---

[6] This implication is cancelable, e.g., 'Had Bob not come to the party, Karen would not have come. In fact, Karen would not have come under any circumstance.'

[7] This mirrors a conclusion drawn by Anna Wierzbicka (1997):

> English does have at least one specific frame to mark "true counterfactuals" (in Barwise's sense): if there may be some doubt about the affirmative "*if* + pluperfect + *would*"-sentences, the negative ones (*if* + pluperfect + Neg. + *would*) clearly do signal semantic "counterfactuality".

Though I disagree that the counterfactuality of such constructions is a semantic phenomenon, I maintain that such constructions are invariably understood as counterfactual because of the pragmatic implication that they carry.

> (33) Wittgenstein had not yet met Russell.

(33) carries with it the implication that Wittgenstein would meet Russell at some later date. It would clearly be misleading and unhelpful to assert the following:

> (34) Wittgenstein had not yet invented the microwave oven.

This is because we do not believe that Wittgenstein *ever* invented the microwave oven. Presumably, there were an infinite number of things that Wittgenstein had not done at the time in question, so we would need a good reason for actually mentioning any of them in conversation. One such reason is that the actions in question are ones that we might have expected him to have done but that he actually had not (or, at least, not at the time in question). However, the use of the *simple past* with 'not' does not seem to carry this implication with it, e.g.,

> (35) Wittgenstein did not invent the microwave oven.

is highly assertable (or, at least, more assertable than (30)). Of course, (30) could become assertable if its implicature were cancelled, as in the following:

> (36) Wittgenstein had not invented the microwave oven by 1912, and never did so.

This provides strong evidence that the past perfect + 'not' implicates counterfactuality in whatever context it occurs.

This explains one difference between (28) and (29). (29) carries the pragmatic implication that its antecedent is contrary to fact, while (28) does not. The use of the simple past with 'not', as in the protasis, does not carry with it the implication of counterfactuality that the past perfect does when combined with 'not'. For instance, (37) does not carry with it the pragmatic implication that Bobby took the cake out of the oven, but (38) does.

(37) If Bobby did not take the cake out of the oven, it burnt.
(38) If Bobby had not taken the cake out of the oven, it would have burnt.

Similarly, (28) does not carry with it the implication that Oswald killed Kennedy, while (29) does.

This, by itself, is not enough to explain why (28) is assertable while (29) is not. Let us therefore examine (28) further. Interpreting (28) as (28a), we see that someone who asserts (28) implicates he has non-truth functional grounds for thinking some past time satisfies (28b) with respect to 't_1'.

(28b) If ~(Oswald kills$_{t1}$ Kennedy), then [There is a t_2: $t_1 < t^*$] (Ruby kills$_{t2}$ Kennedy)

(28b) in no way implies that t_1 is earlier than t_2. Therefore, unlike (29), the antecedent of (28) would not usually be taken to be of a form which, if asserted, would implicate that the speaker thought the event described in the protasis occurred earlier than the event described in the apodosis. Therefore, one could not infer from an assertion of this conditional that the speaker believes that there are grounds for thinking that a process was in place that would lead to Ruby's killing Kennedy in the event that Oswald failed. Such a process would require the event described by the apodosis to occur at a later time than the event described in the protasis. Therefore, by treating (28) as (28a), we do not get the implication that Ruby, for instance, had instructions to kill Kennedy if Oswald failed.

The use of the simple past in the protasis of (28), rather than the past perfect, will suggest that Oswald not killing Kennedy is to be taken to be an epistemic possibility, at least for the moment (rather than a contrary-to-fact hypothesis, as for (29)). It works in a way similar to the simple past verb in the protasis of (37) which invites us to consider the real possibility that Bobby did not take the cake out of the oven while not implicating any contrary-to-fact claim about what would have happened if Bobby had not taken the cake out of the oven even in the circumstance that he did. In the case of (28), since Kennedy's death will be taken to be established, the speaker implicates he has grounds for thinking either Oswald or Ruby killed Kennedy. This is a reasonable thing to assert because, according to my story, Oswald and Ruby are the only possible killers. In the case of (23),

(23) If Oswald didn't kill Kennedy, someone else did [kill Kennedy].

the sum of the implication is that Kennedy was killed. This is really just a logical consequence of (23). The point of asserting (23) could only be to suggest that while there is no doubt about Kennedy's being killed, there may be some about Oswald's having done it.

Summary

My proposal has three important advantages over standard possible worlds theories of 'counterfactual' conditionals. First, it does not introduce any metaphysically dubious entities such as possible worlds. Second, it allows us to treat the verbs that appear in conditional statements in exactly the same way we would treat them elsewhere, resulting in a pleasing economy of theory and, most importantly, preserving "semantic innocence." Third, it forms a unified account of indicative and backshifted conditionals. On my account, backshifted conditionals are treated as tensed indicative conditionals and similar considerations contribute to the assertability of backshifted conditionals as to other indicative conditionals.

CHAPTER 8
BUSINESS ARISING

An important point of interest for any logic is which theorems it finds to be valid. As I noted in Chapter 3, most philosophers and logicians argue that, as far as so-called counterfactual or subjunctive conditionals go, there seems to be counterexamples to each of the following theorems.

(1) Transitivity: $((\phi > \iota) \wedge (\Gamma > \phi)) \rightarrow (\Gamma > \iota)$
(2) Contraposition: $(\phi > -\iota) \rightarrow (\iota > -\phi)$
(3) Strengthening antecedents: $(\phi > \iota) \rightarrow ((\phi \wedge \Gamma) > \iota)$

I am not interested here in the properties of formal languages, but it is worthwhile to explore the use of the supposedly corresponding argument forms in natural language discourse.

Each of these theorems contains the counterfactual operator '>'. According to my account of backshifted conditionals, the motivation behind the introduction of a counterfactual operator is misplaced. So-called counterfactual conditionals are merely a kind of quantified material conditional, so there is no need to introduce a special sort of operator in order to represent their logical form. Since I do not employ a counterfactual operator in my logic for these conditionals, I will not be interested in analyzing (1)-(3). Instead, I will look at some ordinary language examples that are supposed to instantiate these theorems and endeavor to explain our usual reaction to them.

The interest of this lies in the threat the claim about the supposedly corresponding natural language arguments are invalid poses for my analysis of backshifted conditionals. For, as we will see, on my account, all of the corresponding arguments are in fact semantically valid. Thus, we need to examine the relevant argument forms and decide whether the data constitutes evidence against my account.

The following three sentences are offered as counterexamples to transitivity for "counterfactual" conditionals:

> (4) If Carter had not lost the election in 1980, Reagan would not have been president in 1981.
> (5) If Carter had died in 1979, he would not have lost the election in 1980.
> (6) If Carter had died in 1979, Reagan would not have been President in 1981.

According to the orthodox view, if transitivity is valid for so-called counterfactual conditionals, (6) should follow from (4) and (5). However, (4) and (5) are both true while (6) is false, so (6) does not follow, and this must be an invalid argument.

On my account, the above three sentences are analyzed as follows:

> (4a) [There is a t_1: $t_1 < t^*$] (if ~[There is a t_2: $t_2 < t_1$ and t_2 is in 1980] (Carter loses$_{t2}$ the election), then [There is a t_3: $t_3 > t_1$]~[There is a t_4: $t_4 < t_3$ and t4 is in 1981] (Reagan is$_{t4}$ President))
> (5a) [There is a t_1: $t_1 < t^*$] (if [There is a t_2: $t_2 < t_1$ and t_2 is in 1979] (Carter dies$_{t2}$), then [There is a t_3: $t_3 > t_1$]~[There is a t_4: $t_4 < t_3$ and t4 is in 1980] (Carter loses$_{t4}$ the election))
> (6a) [There is a t_1: $t_1 < t^*$] (if [There is a t_2: $t_2 < t_1$ and t_2 is in 1979] (Carter dies$_{t2}$), then [There is a t_3: $t_3 > t_1$]~[There is a t_4: $t_4 < t_3$ and t_4 is in 1981] (Reagan is$_{t4}$ President))

When analyzed this way, the three sentences are obviously not a counterexample for transitivity. The first thing to say is that (4a)–(6a) do not have the form of conditionals, so they are not candidates for the analogy for (1). They all have a quantifier outside the scope of the connective. Even if we ignore this fact, (4a)-(6a) still cannot be a counterexample. To be such a counterexample, the antecedent of (4a) would have to be the consequent of (5a). Moreover, the antecedent embedded in the quantifier of (6a) is false for every time, since there was no time in 1979 at which Carter died. So, in addition to the first reason, this cannot be a counterexample because the conclusion is trivially true. We can also discover (6a) to be trivially true on the grounds that the first quantifier in (6a) allows time t_1 to be any time earlier than t* (the time of utterance), and even a time before 1979, in which case there would be no t_2 in 1979 that is earlier than t_1. Of course, the t_1 that any speaker who utters (4) has in mind is not earlier than 1979, but that is implicated rather than implied. An interlocutor understands that the moment in question must be no earlier than 1979, otherwise an utterance of the conditional makes little sense under any context.

The counterexample to transitivity appears not to be a counterexample at all according to my account. However, we are looking in the wrong place. I have argued throughout this book that we are just as often interested in the conversational implicature of an utterance of a sentence as its truth or falsity. We must be sensitive to information conveyed by utterances of sentences under normal circumstances, and we cannot always describe this information by considering austere meaning alone. We also need to consider the context of the utterance, background assumptions and other pragmatic considerations. These considerations do not only apply to individual sentences, but also to arguments. When evaluating an argument, we do not only consider the relation between the truth of the premises and the truth of the conclusion, but also relations between what utterances of the premises and utterances of the conclusion standardly implicate.

Possible worlds and entailment-based theories tell us something useful about the way we think about certain kinds of conditionals. However, I have argued that they are mistaken to assume that what they tell us has anything to do with the semantics of the expressions they purport to analyze. Rather, these theories are best seen as keeping track of pragmatic considerations that arise from the utterance of the expressions under consideration. When examining the standard pragmatic implicatures of utterances such as (4), (5) and (6) we can expect certain parallels with possible worlds and entailment-based theories.

Let us consider each of the sentences in the above counterexample. We cannot explain our favorable reaction to (4) without considering what an utterance of that sentence standardly implicates. If a speaker knew that Carter lost in 1980 but had no idea that Reagan was the Republican candidate that year, we would not expect her to utter a conditional like (4), even though she could know the conditional to be true on the grounds that the antecedent of the conditional embedded in (4a) is false. We understand that a speaker would not go to the trouble of uttering a conditional like (4) unless she wanted to convey the idea of some sort of connection between the events described in its protasis and apodosis. If all she knew was that Carter lost, we would expect her to simply have said ‘Carter lost the election in 1980’ and have been done with it. If she uttered the conditional (4), she would violate Grice’s Maxim of Quantity by providing less information than would ordinarily be deemed appropriate. We must therefore interpret an utterance of (4) as an attempt to convey the idea of a connection between Carter not losing in 1980 and Reagan not being president in 1981. We react favorably to (4) because we believe there is such a connection. The US constitution ensures that people who run for president and who do not lose generally become President shortly thereafter.

The antecedent of the conditional embedded in the quantifier of (4a) could be satisfied if Carter chose to withdraw from the 1980 presidential race or even if he died before the election happened. However, an utterance of (4) would not countenance such a possibility under normal circumstances. There are a number of reasons for this. Since it is common knowledge that Carter ran in the 1980 election, background knowledge is probably enough for us to infer that, by talking of Carter not losing the election, the speaker intends to say something about what might have transpired following a Carter victory, not what might have transpired after a Carter withdrawal. Moreover, in normal contexts, an utterance of (4) just does not make sense without the assumption that Carter actually ran for president because there is no apparent connection between a person not running for president and him/her being president the following year. If an utterance of (4) were to make sense without the assumption that Carter ran for president, it would have to occur in the context of a story explaining the link between Carter not running and Reagan not being president.

We can also infer that the speaker believes that Carter lost from her use of the past perfect together with 'not' ('had not lost'). I argued in Chapter 7 that utterances of sentences with a past perfect verb together with 'not' standardly implicate a counterfactual hypothesis. Therefore, an utterance of (4) standardly implicates that Carter lost the election in 1980 while speculating about what might have happened in the contrary-to-fact case that he did win. The speaker gives no indication that she wants to consider a situation even further removed from what actually transpired, such as one in which Carter never ran at all. If the speaker really did want to indicate that she was considering such a circumstance, we would ordinarily expect her to give us a clue that this is what she intended.

This sort of pragmatic consideration is mirrored by the idea of "similarity" between possible worlds employed in Lewis's analysis. Lewis argues that a counterfactual is true just in case its consequent is true in those worlds closest to the actual world in which its antecedent is true. The stipulation that we consider only those worlds closest to the actual world ensures that we do not consider counterfactual situations in which background conditions and assumptions are drastically changed. For example, on Lewis's account, a world in which Carter did not run in the 1980 election is presumably less similar to the actual world than a world in which he ran and won. Therefore, those worlds at which it is true that Carter did not lose in 1980 because he did not run and where it is false that he was President in 1981 will be less similar to the actual world than those worlds where Carter did not lose because he ran and won and where it is true that he was President in 1981. I have argued that Lewis is mistaken to claim that these considerations about possible worlds tell us anything about the truth conditions or the semantics of these conditionals. Given the nature of the tense expressions that occur in these conditionals, we have little reason to believe that they are literally *about* possible worlds. Thinking about possible worlds in the way that Lewis suggests could certainly be a useful heuristic device for deciding whether a conditional is properly assertable, but it is no more than that.

There are contexts in which a speaker could assert (4) without implicating the assumption that, in the hypothesis described in the protasis, Carter actually ran in 1980. The story given in (7), provides such a context:

> (7) It was September 1998. Brad was recovering slowly from his recent bout of amnesia, and his memory was still patchy. Today, he was trying to piece together his memory of past American presidents. He remembered a few of them, but he could not recall who had been president in 1981. Brad had always been a Democrat and a huge fan of Jimmy Carter, so he remembered with some sadness that Carter had run against Ronald Reagan one year and that his favorite politician had suffered his only loss in a presidential election. Unfortunately, Brad could not remember whether Carter had lost the election in 1984 or in 1980, though he was sure that Walter Mondale had lost for the Democrats one of those two years, so he knew that Carter could not have been the losing candidate both times. He also remembered that Reagan had been president for eight whole years following his victory over Carter and that those eight years had been the only ones during which Reagan had been president.
>
> Brad concentrated hard, trying to remember the year that Carter had lost the election. He was becoming more and more certain that Carter had lost in 1984, not 1980, and if he was right, he had a small clue about who was president in 1981. *If Carter had not lost the election in 1980, Reagan would not have been president in 1981.*

In this story, the protasis of (4), as it occurs in the final sentence, obviously is not intended to suggest a situation in which Carter did not lose in 1980 because he ran and won, but a situation in which Carter did not lose in 1980 because he lost in 1984 instead. The story undercuts the implicature that Carter ran. It also undercuts the implicature that Reagan was President in 1981, and so clearly undercuts whatever else is implicated by more typical uses of (4), something to the effect Reagan became President in 1980 because the only serious rival was Carter. Thus, the case shows that whatever is going on is pragmatic and not semantic.

The fact that the implicature in (4) that Carter ran does not appear in certain contexts should give us good reason to believe that the sentence is not always assertible and that its assertibility does not simply depend on its truth or falsity. For example, a fully informed opponent of the Reagan presidency would never argue as follows:

> (8) I wish Carter had never run in the 1980 election, because if he had never run, he never would have lost it. Moreover, *if Carter had not lost the election in 1980, Reagan would not have been president in 1981.*

Lewis would undoubtedly argue that (4) is not assertable in the context of (8) because the notion of similarity between possible worlds is context dependent. When considered in the context of (8), Lewis would have to say that what counts as the closest possible worlds in which the antecedent of (4) is true somehow change because of the shift in context. However, this move is *ad hoc*. Once we know the relevant facts about the world (and the relevant facts about possible worlds, if there are any), if we know what a sentence means, and we know who uttered it and when, we should have enough information to determine the truth or falsity of that sentence. The claim that a sentence could be discovered to be false simply by discovering that the speaker uttered something else before she uttered that sentence is counterintuitive unless that sentence referred in some way to what the speaker had said before, which (4) does not when uttered in the context of (8). This sort of context-dependency offers us a good clue that the phenomenon we are interested in is pragmatic, not semantic, in nature. Admittedly, this is not a conclusive argument. It is easy to imagine someone urging that there is a hidden demonstrative reference of some kind to a similarity relation the speaker has in mind or some complex rule that takes one from total context to a relevant similarity relation. However, in the absence of a concrete proposal, there is not much I can say in response to such an idea. However, what I can say is that I have an account of the phenomena here in terms of a unified semantics for tensed expressions and an independently well-grounded theory of implicatures, and my opponent has at best a sketch of a proposal with no idea how to fill in the details.

To make the implicature in (4) explicit, the speaker could have said something like (9):

> (9) If Carter had not lost the 1980 election after running in that election, Reagan would not have been president in 1981.

Under normal circumstances, a person would be prepared to utter (4) just in case she was prepared to utter (9). This suggests that utterances of (4) and (9) convey the same information under such circumstances. By "normal circumstances" I mean circumstances that assume certain background assumptions. One could imagine scenarios in which, e.g., Anderson (the independent candidate) was leading Carter by 30 percentage points, and the issue was whether it was going to be Anderson or Reagan who wins the election. However, what we know about the 1980 election, an utterance of (4) would implicate that the speaker is also prepared to utter (9).

An utterance of (5) considers a situation in which Carter died in 1979. Nobody would utter this conditional simply on the grounds that he did not believe that Carter died in 1979. We interpret someone who utters (5) as suggesting a relation between the events described in its protasis and apodosis. We react favorably to (5) because we believe there is such a relation. A person who dies cannot lose an election the following year losing an election requires that you run in it.

Why do not we accept (6) after accepting (4) and (5)? The problem is that utterances of (4) and (5) are typically made against different background assumptions.

An utterance of (4) implicates that Carter actually ran for president in the counterfactual situation that is under consideration. The protasis of (5) implies that this assumption is false. The reason we do not believe (6) follows (4) and (5) is similar to the reason we do not believe that (6) follows from (9) and (5):

> (9) If Carter had not lost the election in 1980 after running in that election, Reagan would not have been president in 1981.
> (5) If Carter had died in 1979, he would not have lost the election in 1980.
> (6) If Carter had died in 1979, Reagan would not have been President in 1981.

(6) fails to follow from (5) and (8) because (5) does not guarantee that the second conjunct of the (embedded) conditional in (8) is true for a suitable past time. The difference between (8)-(6) and the original argument is that the assumption about Carter actually running in the election is only implicated by the protasis of (4), rather than explicitly stated, as it is in the protasis of (8). Nonetheless, it is clear that our commitment to asserting (4) in the context depends on the assumption that Carter ran, and so we will not be committed to Reagan's not having been president in 1981 on the assumption that Carter did not run, even though that means that he did not lose the election.

It is interesting to note that counterexamples very similar to the above can also be given using conditionals that most possible worlds theorists would describe as "indicative." Consider (10), (11) and (12):

> (10) If Bush does not lose to McCain, Bush will run for President.
> (11) If Bush beats everyone except Keyes, he will not lose to McCain.
> (12) If Bush beats everyone except Keyes, Bush will run for President.

An utterance of (10) standardly implicates that other candidates like Keyes are so far behind that only McCain and and Bush stand a chance of winning the Republican nomination. However, the protasis of (11) implies that this assumption is false. Though assertions of (10) and (11) would be perfectly acceptable by themselves, we would not be prepared to assert (12) after asserting (10) and (11). The fact that the very same phenomena can occur with so-called indicative conditionals suggests either that most philosophers have drawn the line between "indicative" and "subjunctive" conditionals incorrectly, or, what is more likely, that the phenomenon in question is not due to the conditionals' (supposed) mood at all. If the phenomenon, which is supposed to show something special needs to be said about "counterfactuals," can be shown to be attached to examples outside the class, we neutralize the evidence for the existence of that class.

Let us next consider a counterexample to Contraposition. A possible worlds theorist would argue that if contraposition held for "counterfactual" conditionals, (14) would follow from (13):

> (13) If it were to rain heavily at noon, the farmer would not irrigate his field at noon.
> (14) If the farmer were to irrigate his field at noon, it would not rain heavily at noon.

However, it is argued, whenever (13) is true, (14) appears to be false. Indeed, (14) appears to be false under any circumstances. Therefore, contraposition does not hold for counterfactuals.

An analysis of conditionals that pays attention to their use of tense expressions gives us a useful explanation of why this argument appears counterintuitive. (13) and (14) are to be analyzed as (13a) and (14a) respectively:

(13a) [There is a t_1: $t_1 < t^*$ and t_1 is at noon] (If it rains$_{t1}$, then ~[there is a t_2: $t_2 > t_1$ and t_2 is at noon] (the farmer irrigates$_{t2}$ his field))
(14a) [There is a t_1: $t_1 < t^*$ and t_1 is at noon] (If the farmer irrigates$_{t1}$ his field, then ~[there is a t_2: $t_2 > t_1$ and t_2 is at noon] (it rains$_{t2}$))

The analysis of (13) and (14) offered in (13a) and (14a) might at first seem peculiar. The relativized future models 'would irrigate' and 'would rain' in (13) and (14) force an analysis of each of the two sentences as saying that there is a time t_2, later time than another time t_1, but such that t_1 and t_2 are both at noon. How can we make sense of this? One could argue that the expression 'noon', as it is used here, refers to a *period* of time surrounding midday, not the *precise* moment of 12pm Then it would be possible for one time 'at noon' to be later than another time 'at noon' (say, for instance, 12.10 and 12.01pm). However, this does not provide a good general strategy for dealing with these sorts of cases because expressions like '12pm precisely' could easily be substituted for 'noon' in (13) and (14), and both sentences would still make sense.

We can perhaps take comfort in the fact that an utterance of (13) sounds a little odd. We would ordinarily prefer something like (15):

(15) If it were to rain heavily at noon, the farmer would not irrigate his field.

Clearly, (15) does not give us the same problem as (13) because there is no adverbial restricting the time picked out by the verb in the apodosis. However, it appears that the oddness of (13) has more to do with the clumsiness of the repetition of 'at noon' in the apodosis than with any semantic problem in the sentence. For instance,

(16) If it were to rain heavily at noon, the farmer would not irrigate his field then,

seems like a perfectly normal thing to say, though the indexical ‘then’ in the apodosis clearly restricts the time of the absence of irrigation to noon.

However we proceed, we need to look elsewhere than the truth conditions of (13) and (14) in order to interpret utterances of these sentences properly. (13) is made trivially true by there undoubtedly being some time before the time of utterance at which it does not rain, yet nobody would utter it on those grounds alone (similarly, (14) is trivially true because there is undoubtedly some time before the time of utterance at which the farmer did not irrigate his field). Therefore, we need not be unduly concerned that, in reality, there cannot be two times, both at noon, such that one time is later than the other. Since both (13) and (14) are trivially true anyway, the impossibility of two such times will have no significant effect on the truth or falsity of the sentences. We are already forced to look at pragmatic considerations in order to interpret the sentences properly.

Another peculiarity of the analyses in (13a) and (14a) is that they suppose that we can use past tense verbs to talk about hypothetical events in the future. This phenomenon was explained in the previous chapter. It would be perfectly appropriate to assert (13) when discussing the question of whether the farmer will irrigate his field at noon tomorrow, even though the protasis contains the tense expression ‘were to rain,’ which, as I argued in Chapter 7, is an equivalent of the simple past ‘rained’. As I also explained in Chapter 7, we use the past tense in this way in order to give the sentence a “universal” import. In the case of (13), the universal reading derived from the sentence is similar in import to ‘whenever it rains at noon the farmer tends not to irrigate his field at noon’. The reason why one would assert (13) rather than

(17) If it rains heavily at noon, the farmer will not irrigate his field at noon,

is that one wants to indicate some necessary, regular or lawlike connection between it raining and the farmer not irrigating his field. The reason why there is a difference between (13) and (17) in this regard is because, according to my analysis of (13), a quantifier binds across the conditional, but this is not so for (17), which is analysed as in (17a):

> (17a) If [there is a t_1: $t_1 = t^*$ and t_1 is at noon] (it rains_{t1}), then ~[there is a t_2: $t_2 > t^*$ and t2 is at noon] (the farmer irrigates_{t2} his field))

Ordinarily, when we say that one thing will happen later than another, we have some good reason for doing so. For example, we might believe that one event, if it occurs, will be the cause of another event so that the first event must precede the second. An utterance of (13) would standardly implicate that rain at noon would cause the farmer not to irrigate his field at noon (and not vice versa). It also implicates that it would be possible for the farmer to irrigate his field if he chose to do so. For instance, you would not utter a sentence like (13) if you believed that the farmer was lying on his deathbed, unable to get up and do anything, let alone irrigate his field. In such a situation, the farmer would not irrigate his field at noon, but the rain would not be the cause of him not doing so.

Let us assume that it rains at noon. Clearly, the farmer will not decide right at noon whether he is going to irrigate his field then. If he is not irrigating his field at noon because of rain, he must have made some decision beforehand, maybe on the basis of the weather forecast, or maybe on the grounds that it was raining before noon and it did not appear to be letting up, or maybe because there were storm clouds gathering shortly before noon. He may have made his decision at the last possible minute, e.g., he could have been ready to turn on the irrigation pumps at noon when a sudden downpour make him think better of it. Whatever story we give, the farmer must have had knowledge that it was going to rain before deciding not to irrigate at noon.

This seems to be what is implicated by the temporal ordering of the two events in the analysis (13a); there was some time earlier than noon when the farmer decided not to irrigate his field at noon because he believed it would rain at noon. Therefore, it is the rain at noon that causes the farmer not to irrigate his field, but indirectly so, because the impending rain caused the decision to be made at some time before noon.

This should make clear why (14) would not be an appropriate sentence to assert. We cannot imagine the farmer not irrigating his field at noon being the cause of it not raining at noon, whether directly or indirectly. An utterance of (14) implicates such a connection, and it therefore does not make sense. This is why we react unfavorably to an utterance of (14) while, at the same time, reacting favorably to an utterance of (13).

As in the case of the counterexample to transitivity, it is not difficult to find an "indicative" equivalent of (14) that sounds as odd as (14) itself. For instance, we would assert (17) but not (18):

(17) If it rains at noon, then the farmer will not irrigate his field at noon.
(18) If the farmer irrigates his field at noon, then it will not rain at noon.

A similar causal connection between the events described in the protasis and the events described in the apodosis is suggested here too, though the absence of the past-tensed verbs remove the universal import of utterances of (13) and (14).

However, the phenomenon does not occur when all the verbs in the conditionals are simple past tense. For example, both (19) and (20) would be perfectly reasonable things to utter under the appropriate circumstances.

(19) If it rained at noon, the farmer did not irrigate his field at noon.
(20) If it the farmer irrigated his field at noon, it did not rain at noon.

Suppose that you have forgotten whether it rained at noon yesterday, but you know that the farmer irrigated his field then. Then, the information conveyed by an utterance of (20) would be helpful to you when trying to remember what the weather was like. What makes the difference here is that (19) and (20) contain no future tensed verbs or relativized future modals. Nothing in (19) and (20) requires the events described in the protasis to be earlier or later than the events described in the apodosis. Thus, we interpret utterances of (19) and (20) as indicating something about epistemic possibilities rather than causal connections. This strongly suggests that the phenomenon that occurs in the alleged counterexample to Contraposition is due to tense of the verbs rather than their (supposed) mood.

Finally, let us consider Strengthening Antecedents:

(21) If Otto had come, it would have been a lively party.
(22) But if both Otto and Anna had come it would not have been a lively party.

The traditional argument says that if Strengthening Antecedents were valid for counterfactual conditionals, (21) and (22) could not be true together. However, we can perfectly imagine how they might be true together, so strengthening Antecedents cannot be valid for counterfactual conditionals.

Again, an analysis of backshifted conditionals that is sensitive to pragmatic considerations provides a good explanation of why a person could utter both (21) and (22). An utterance of (21) implicates some sort of connection between Otto coming to the party and the party being a lively one, e.g., because Otto is a lively fellow who becomes the life and soul of every party that he attends. However, just as an utterance of

> (4) If Carter had not lost the election in 1980, Reagan would not have been president in 1981.

standardly implicates an assumption that Carter did not suffer the misfortune of dying before the election, an utterance of (21) standardly implicates an assumption that, should Otto have come to the party, nothing untoward or unexpected would have happened to him or to the other guests at the party. This implicature can be cancelled, as in (23):

> (23) If Otto had come, it would have been a lively party. But not because of Otto's presence. His grandmother was run over by a bus this morning, and he's completely distraught. It would have been a lively party despite Otto.

This shows that the implicature generated by (21) assumes some appropriate background, since no one could suppose that Otto's coming under any possible circumstances would ensure that the party was lively. Suppose Otto hates Anna and he is always in a bad mood when she is around. Given this information, we can cancel the implicature of (21) as in (24):

> (24) If Otto had come, it would have been a lively party. But not because of Otto. Anna turned up there, and you know how she always puts him in a terrible mood! It would have been a lively party despite Otto.

Given the right sort of information about Otto and Anna, it is easy to see how both (21) and (22) could both be assertable together. The additional sentences in (22) simply cancel the implicature in (21), relative to appropriate background knowledge, that nothing untoward or unexpected would have happened to upset Otto's usual cheerful self if he had come to the party. Again, we can see a parallel here with the notion of similarity used in possible worlds theories. The worlds in which Otto comes to the party that are most similar to the actual world are worlds in which Anna did not come to the party. This suggest that what guides the possible worlds theorist in deciding what worlds are to be said to be most similar to the actual world is whether what is true in those worlds is consistent with what is standardly implicated by an assertion of the conditional under analysis.

Again, we can create a similar sort of example using sentences that would ordinarily be considered indicative:

(25) If Otto comes to the party, it will be lively.
(26) If Otto and Anna come to the party, it will not be lively.

This again suggests that it is not the mood of their verbs that causes these sentences to behave in the way they do. However, in this case, it does not appear that the phenomenon is due to tense either. (27) and (28) could be assertable under appropriate circumstances, yet neither contains a future tense verb or relativized future modal:

(27) If Otto came to the party, it was lively.
(28) If Otto and Anna came to the party, it was not lively.

What is going on is nothing more than a standard implicature being cancelled out by a change to the protasis. An account sensitive to the pragmatic considerations involved in the use of these conditionals thus appears particularly fruitful.

I have considered three alleged counterexamples to certain forms of argument containing backshifted conditionals. My analysis of backshifted conditionals does not allow these to be counterexamples in the traditional sense, since they all have true conclusions. However, my analysis does adequately explain why we react to these arguments in the way we do in such a way that appears subtler and more sensitive to the nuances of the English language than more traditional analyses. This lends support to my claim that my analysis of these conditionals is the right one. It is of particular importance to notice that the phenomenon appealed to in all three cases also occurs in the case of conditionals that are acknowledged to be "indicative." This demonstrates that this kind of data cannot be used to show that backshifted conditionals require a special semantic treatment different from that of indicative conditionals. No one thinks the possible worlds approach appropriate for the indicative cases, and almost no one, anymore, suggests we need a special connective in these cases. So by and large it is accepted that the explanation here must be pragmatic. But then clearly where the same phenomenon shows up in the case of backshifted conditionals, we should expect it to receive a similar explanation. Moreover, as I have shown, we can straightforwardly do so once we take seriously the tense structure of these conditionals.

CHAPTER 9
CONCLUSION

In this dissertation, I have offered an account of so-called "subjunctive" conditionals that gives full consideration to their use of tense expressions. I have presented this account in the context of a truth-conditional semantics for English, which, I believe, offers a particularly illuminating background for considering the semantics and pragmatics of tense expressions in English and other natural languages.

The primary motivation behind my account of so-called "subjunctive" conditionals is the preservation of semantic innocence. Wherever possible, we should avoid saying that expressions are radically ambiguous. To me, it seems outrageous to suppose that a mere change of tense could signal a radical difference in the truth conditions of any kind of statement, as most traditional accounts of these conditionals do. Moreover, the very same verb constructions that appear in so-called "subjunctive" conditionals are treated elsewhere as tense expressions. My account treats the tense expressions in conditionals exactly as they appear to be. This consideration alone ought to make such an account worthy of consideration. Why, therefore, have so few philosophers given any thought to treating conditionals in this way?

The first major problem is that almost the entire literature on conditionals assumes both the traditional distinction between indicative and counterfactual/subjunctive conditionals and the traditional classification of a conditional's antecedent and consequent. In Chapter 2, I presented some very good reasons for thinking that these assumptions are wrong. The falsity of a conditional's antecedent does not give us grounds for putting it in a special semantic category, and the grammatical distinction between indicative and subjunctive simply does not exist in English to the extent that most philosophers and logicians suppose.

The phenomenon of the questionable yet influential assumption is common in the history of philosophy. We see philosopher X present an influential and persuasive new theory, but certain unquestioned assumptions are smuggled in with it unnoticed. A debate begins. Some support X while others argue against him. However, the terms of the debate were set when X made his original assumptions, and we become so caught up in the finer points of X's theory and its rivals that we forget to question the assumptions on which the whole debate rests. In our case, X could be any one of Stalnaker, Lewis, Goodman or Chisholm.

A second problem is that traditional theories of so-called "subjunctive" conditionals confuse pragmatic considerations with semantic ones. They confuse those factors that contribute to the felicity of a conditional's utterance with the truth conditions of that conditional. This is an understandable mistake, and similar ones have been made in other areas of philosophy. H.P. Grice (1989) gives a number of examples of this sort of phenomenon.

In Chapters 2 and 3, I discussed the theories of Goodman, Chisholm, Mackie, Stalnaker and Lewis in some detail and highlighted some major internal problems with these theories. My semantics of so-called "subjunctive" conditionals avoids the problems encountered by these other accounts popular in the literature. Possible worlds-based theories suffer because they employ a problematic "similarity" relation and entailment-based theories fail because they cannot be stated without circularity. Since my semantics employs no similarity relation, and since it does not involve specifying “relevant conditions” for conditionals, as entailment-based theories do, it bypasses these problems completely.

Though it is not my primary concern here, an added benefit of my treatment of so-called “subjunctive” conditionals is that it avoids an ontology of possible worlds. I have not attempted to argue against the legitimacy of possible worlds as a philosophical device. However, economy of hypothesis ought to be an important consideration in the evaluation of any philosophical or scientific theory (and that includes semantic theories). My theory achieves great economy of hypothesis by avoiding talk of possible worlds altogether. Some semanticists seek to avoid the more controversial aspects of an ontology of possible worlds by treating them as sets of functions from sentences (or propositions) to truth values rather than as metaphysical entities (e.g., Nute, 1980). However, such philosophers who seek to use possible worlds to specify the meaning of English expressions are left with the somewhat counterintuitive claim that sentences of English are literally about sets of functions. All things being equal, a theory of conditionals that avoids talk of possible worlds altogether must be preferable.

Treating conditionals in the way that I suggest will have no significant impact on how we employ conditionals in philosophical theories. What this dissertation argues is not that the import of utterances of conditionals is different from what we thought it was, but that it depends more on pragmatic considerations than most popular theories suppose. The main lesson to be learned is that many of the conditionals that we use to state our theories in ethics, philosophy of science and other areas rely significantly on pragmatic considerations to achieve their desired effect.

REFERENCES

Adams, Ernest W. 1970. "Subjunctive and Indicative Conditionals," In *Foundations of Language* 6, pp. 89-94.

–1975: *The Logic of Conditionals.* Dordrecht, The Netherlands: Reidel.

Athanasiadou, A., and Dirven, R., eds., 1997. *On Conditionals Again: Current Issues In Linguistic Theory*, 143. Philadelphia, PA: Benjamins.

Ayers, M. R. 1965. "Counterfactuals and Subjunctive Conditionals." *Mind*, 74, pp. 347-364.

Bennett, Jonathan 1988. "Farewell to the Phlogiston Theory of Conditionals." *Mind*, 97, pp. 509-527.

---1995. "Classifying Conditionals: The Traditional Way is Right." *Mind*, 104, pp. 331-334.

Blue, N. A. 1981. "A Metalinguistic Interpretation of Counterfactual Conditionals." *Journal of Philosophical Logic* 10, pp. 179 – 200.

Chisholm, Roderick M. 1946. "The Contrary-to-Fact Conditional." *Mind*, 55, pp. 215-235.

Dahl, O. 1997. "The relation between past time reference and counterfactuality: a new look." in Athanasiadou and Dirven, eds., *On Conditionals Again: Current Issues In Linguistic Theory*, 143. Philadelphia, PA: Benjamins.

Davidson, D. 1984. *Inquiries into Truth and Interpretation.* Oxford: Oxford University Press.

Davies, Eirian C. 1979. *On the Semantics of Syntax: Mood and Condition in English.* Highlands, NJ: Humanities Press.

Davis, Wayne A. 1979. "Indicative and Subjunctive Conditionals." *The Philosophical Review*, 88, pp. 544-564.

Dudman, V. H. 1984. "Parsing 'If' – Sentences." *Analysis*, 44, pp. 145-153.

---1986. “Antecedents and Consequents.” *Theoria*, 52, pp. 168-199.

---1988. “Indicative and Subjunctive.” *Analysis*, 48, pp. 113-122.

---1989. “Vive la Revolution!” *Mind,* 98, pp. 591-603.

---1992. “Probability and Assertion.” *Analysis*, 52, pp. 204-211.

---1994a. “On Conditionals.” *Journal of Philosophy*, 91, pp. 113-128.

---1994b. “Against the Indicative.” *Australasian Journal of Philosophy*, Vol. 72, No. 1, pp. 121-148.

Edgington, Dorothy 1995. “On Conditionals.” *Mind*, 104, pp. 235-329.

Ellis, Brian 1984. “Two Theories of Indicative Conditionals.” *Australasian Journal of Philosophy*, Vol. 62, pp. 52 –70.

Etchemendy, John 1990. *The Concept of Logical Consequence*. Cambridge, MA: Harvard University Press.

Fine, Kit 1975. Review of Lewis’s *Counterfactuals. Mind*, 84, pp. 451-458.

Gabbay, Dov 1976. *Investigations in Modal and Tense Logics with Application to Problems in Philosophy and Linguistics*. Dordrecht, The Netherlands: Reidel.

Goodman, Nelson 1955. *Fact, Fiction and Forecast*. Cambridge, MA: Harvard University Press.

Grice, H. P. 1989. *Studies In The Way of Words*, Cambridge, MA: Harvard University Press.

Harper, W. L., Stalnaker R., and Pearce, C. T., eds., 1981. *Ifs*. Dordrecht, The Netherlands: Reidel.

Jackson, Frank 1977. “A Causal Theory of Counterfactuals.” *Australasian Journal of Philosophy*, 55, pp. 3-21.

---1987. *Conditionals,* Oxford: Blackwell.

---1991. ed. *Conditionals,* Oxford: Blackwell.

James, Francis 1980. *Unified Theory of the English Subjunctive*. Ph.D. Thesis: UC Berkeley.

---1986. *Semantics of the English Subjunctive.* Vancouver: University of British Columbia Press.

Jespersen, Otto 1954-58. *A Modern English Grammar on Historical Principles.* London: Allen & Unwin.

Lepore E. and Ludwig K. 2000. "Outline of a Truth Conditional Semantics for Tense." in Quentin Smith, ed., *Tense Time and Reference.* Oxford: Oxford University Press.

Lewis, David 1973. *Counterfactuals.* Cambridge, MA: Harvard University Press.

Ludwig, Kirk 1997. "The Truth about Moods." in *Protosociology, Cognitive Semantics I: Conceptions of Meaning* 10, pp. 19-66.

Mackie, J. L. 1962. "Counterfactuals And Causal Laws." In R. J. Butler, ed., *Analytical Philosophy.* Oxford: Oxford University Press, pp. 66-80.

Nute, Donald 1980. *Topics in Conditional Logic.* Dordrecht, The Netherlands: Reidel.

---1980. "Conversational Scorekeeping and Conditionals." *Journal of Philosophical Logic* 9, pp. 153-166.

---1981. Introduction, *The Journal of Philosophical Logic* 10, pp. 127-147.

---1984. "Conditional Logic." in *Handbook of Philosophical Logic*, Vol. 2, Dordrecht, The Netherlands: Reidel, pp. 387-439.

Pollock, John L. 1976. *Subjunctive Reasoning*. Dordrecht, The Netherlands: Reidel.

---1981. "A Refined Theory of Counterfactuals." *Journal of Philosophical Logic*, 10, pp. 239-266.

Read, S. 1988. *Relevant Logic*. Oxford: Blackwell.

Searle, John 1979. "Indirect Speech Acts." in *Expression and Meaning: Studies in the Theory of Speech Acts.* Cambridge: Cambridge University Press.

Sellars, W. S. 1975. "Counterfactuals." in E. Sosa, ed., *Causation and Conditionals.* Oxford: Oxford University Press, pp. 126-146.

Stalnaker, Robert A. 1981. "A Theory of Conditionals." in Harper, Stalnaker, and Pearce, eds., *Ifs*. Dordrecht, The Netherlands: Reidel.

Tarski, A., 1956. *Logic, Semantics, Metamathematics.* Oxford: Clarendon Press.

Wierzbicka, Anna 1997. “Conditionals and counterfactuals: conceptual primitives and linguistic universals.” in A. Athanasiadou, and R. Dirven, eds., 1997. *On Conditionals Again: Current Issues In Linguistic Theory*, 143. Philadelphia, PA: Benjamins.

Woods, Michael 1997. *Conditionals.* Oxford: Clarendon Press.

BIOGRAPHICAL SKETCH

John Peoples was born in Twickenham, West London, in 1969. Shortly after completing a Bachelor's degree at the University of Manchester, England, he came to the University of Florida in 1993. He received his M.A. from Florida in 1996 after writing a thesis on Tarski's definition of logical consequence. After completing his Ph.D., he hopes to teach philosophy at the university level. His main philosophical interests are in philosophy of language, philosophy of logic, and analytic metaphysics.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Gregory Ray, Chairman
Associate Professor of Philosophy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Kirk Ludwig, Co-chairman
Associate Professor of Philosophy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John Biro
Professor of Philosophy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jean A. Larson

Jean Larson
Professor of Mathematics

This dissertation was submitted to the Graduate Faculty of the Department of Philosophy in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Winfred M Phillips

May, 2000
Dean, Graduate School